UNIVERSITY OF CALIFORNIA SAN DIEGO

Carotenoid biosynthesis and productivity in diatoms

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Biology

by

Olga Gaidarenko

Committee in charge:

James Golden, Chair
Eric Allen
Steven Briggs
Brian Palenik
Julian Schroeder
Maria Vernet

2018

The Dissertation of Olga Gaidarenko is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

2018

DEDICATION

I would like to dedicate this dissertation to my advisor, Dr. Mark Hildebrand, who was an incredibly dedicated and generous mentor, and a dear friend.

I would also like to dedicate this dissertation to my beloved, wonderful grandparents Victor, Elena, Raisa, and Vladimir, in no particular order. I admire you all tremendously.

Finally, I would like to dedicate this dissertation to my favorite (not so) little (anymore) human, Layla Michelle Muñozmartin, who has been one of my greatest teachers in life. Shine on and always remember to stop and feel the breeze as you grow, kiddo.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First, I cannot thank enough my advisor Dr. Mark Hildebrand who was incredibly dedicated to his students and really generous with his time and energy. His support, enthusiasm, patience, and friendship were indispensable throughout this journey, and I am forever grateful.

I am also very grateful to my doctoral committee members for all their support, guidance, and help. Thank you to Dr. James Golden for stepping in as the chair of the committee in the final few months and providing very helpful feedback and advice, as well as for ongoing support throughout my time in the program. Thank you to Dr. Maria Vernet for all the help, mentorship, advice, and support, and for enabling me to do the photopigment work. Thank you to Dr. Steve Briggs for all the support, guidance, help, discussions, and insights. Thank you to Dr. Brian Palenik for all the insightful questions, feedback, suggestions, and support. Thank you to Dr. Eric Allen for all the helpful fresh perspectives, discussions, suggestions, and support. Thank you to Dr. Julian Schroeder for being willing to join the committee in the last couple of months and being very helpful and flexible.

Thank you to all the members of the Hildebrand laboratory, present and past, for being such a wonderful group to be a part of. Thank you especially to Dr. Sarah Lerch, Dr. Roshan Shrestha, Daniel Yee, Corinne Sathoff, Dr. Jesse Traller, and Dr. Eva Sanchez for all the support, especially in the last few months leading up to my defense. Thank you also to Dr. Raffaela Abbriano, Dr. Aubrey Davis, Dr. Orna Cook, Dr. Sarah Smith, Dylan Mills, and Dr. Kalpana Manandhar-Shrestha for all the help, discussions, and being great lab mates, and thank you Maitreyi Nagarkar for being a great honorary lab member.

Thank you also to the Hubbs community and SIO community in general for being such a wonderful part of my experience here and providing a supportive and intellectually stimulating environment. Thank you as well to the members of the Biological Sciences department,

especially Dr. Amy Pasquinelli, Dr. James Posakony, Dr. Don Helinski, and Dr. Kees Murre for mentorship and being so supportive of me. Thank you also to Dr. Susan Golden for encouraging my interest in circadian biology and providing access to her course, which was one of the highlights of my graduate school journey.

Thank you to all the administrative folks who have helped with numerous logistics, especially Annamarie Bryson, Marifel Alfaro, Suzi Harlow, Lien Ngo, Natalie Noles, Cathy Pugh, and Thomas Thomp.

Thank you also to the members of the UC San Diego and Biological Sciences teaching community, especially Dr. Erilynn Heinrichsen – it has been such a pleasure interacting with you, and I learned so much.

Thank you to my family and especially my parents for all the patience, understanding, support, and tremendous sacrifices made so I can be where I am today.

Lots of gratitude also to my wonderful friends who have been so supportive and understanding through this journey, especially Tarsha Stith, Michael Martin, Erica Lo, Nellia Fleurova-Charlton, Duke Allen, Dr. Kelly Beverly, Dr. Eleodora Andreatta, and Katie DiLibero. And, of course, my dearest Layla Michelle Muñozmartin and my sweet kitty-friend CousCous.

Finally, a big thank you to the music community for enriching my life in so many ways and helping me be a stronger, better, and more balanced person.

PUBLISHED/PREPARED FOR PUBLICATION DISSERTATION MATERIAL

Chapter 1, in full, has been submitted for publication. Gaidarenko, Olga; Sathoff, Corinne; Staub, Kenneth; Huesemann, Michael M.; Vernet, Maria; Hildebrand, Mark. “Timing is everything: diel metabolic and physiological changes in the diatom *Cyclotella cryptica* grown in

simulated outdoor conditions." Olga Gaidarenko was the principal author on this paper. We thank Dr. Susan Golden and members of her lab for lending us the ePBR used for this work. We also thank Dr. Daniel Wangpraseurt for assistance with data interpolation. Additionally, we thank Dr. James Golden for providing helpful feedback on the manuscript. This work was supported by U.S. Dept. of Energy grant DE-SC0012556.

Chapter 2, in full, is material currently being prepared for submission for publication. Gaidarenko, Olga; Mills, Dylan W.; Vernet, Maria; Hildebrand, Mark. "Novel *Thalassiosira pseudonana* violaxanthin de-epoxidase-like enzyme (VDL2) catalyzes fucoxanthin biosynthesis." Olga Gaidarenko was the principal researcher and author of this work. We thank Dr. Sarah R. Smith and Dr. Raffaela M. Abbriano for helpful discussions and data that made this work possible, as well as Dr. Roshan P. Shrestha for helpful discussions and providing genetic manipulation tools. Additionally, we thank Dr. Bradley Moore and Dr. Jonathan Chekan for providing valuable perspective and insights. We also thank Dr. James Golden for critically reading the chapter and providing helpful input. This work was supported by U.S. Dept. of Energy grant DE-FOA-0001471.

Chapter 3, in full, is material currently being prepared for submission for publication. Gaidarenko, Olga; Yee, Daniel; Hildebrand, Mark. "Enhanced triacylglycerol (TAG) and protein accumulation in transgenic diatom *Thalassiosira pseudonana* with altered photosynthetic pigmentation." Olga Gaidarenko was the principal researcher and author of this work. We thank Dr. Andrew E. Allen and members of his lab for providing access to the Water PAM used in this work. We also thank Ms. Corinne Sathoff for assistance with cultivation chamber set-up and ImageStream data analysis. Additionally, we thank Dr. James Golden for critically reading this work and providing helpful input. This work was supported by U.S. Dept. of Energy grant DE-FOA-0001471.

VITA

2006 Bachelor of Science, Molecular Biology
University of California San Diego

2018 Doctor of Philosophy, Biology
University of California San Diego

PUBLICATIONS

Gaidarenko O., Yee D., Hildebrand M. Enhanced triacylglycerol (TAG) and protein accumulation in transgenic diatom *Thalassiosira pseudonana* with altered photosynthetic pigmentation. In prep.

Gaidarenko O., Mills D. W., Vernet M., Hildebrand M. Novel *Thalassiosira pseudonana* violaxanthin de-epoxidase-like enzyme (VDL2) catalyzes fucoxanthin biosynthesis. In prep.

Gaidarenko O., Sathoff C., Staub K., Huesemann M. M., Vernet M., Hildebrand M. Timing is everything: diel metabolic and physiological changes in the diatom *Cyclotella cryptica* grown in simulated outdoor conditions. Submitted.

Wangpraseurt D., You S., Azam F., Jacucci G., **Gaidarenko O**., Hildebrand M., Kuhl M., Smith A. G., Davey M. P., Deheyn D. D., Chen S., Vignolini S. Bionic 3D printed corals. Submitted.

Traller, J., Lopez, D., Cokus, S., **Gaidarenko, O.**, Smith, S., McCrow, J., Gallaher, S., Podell, S., Thompson, M., Cook, O., Morcelli, M., Jaroszewicz, A., Allen, E., Allen, A.E., Merchant, S.S., Pellegrini, M., and Hildebrand, M. (2016). Genome and methylome of the oleaginous diatom *Cyclotella cryptica* reveal genetic flexibility toward a high lipid profile. Biotechnol Biofuels 9:258.

Gaidarenko, O. and Xu, Y. (2011). Relationship between regulatory pathways in pluripotent stem cells and human tumors. In Allan, A. (Ed.), *Cancer Stem Cells in Solid Tumors*. Springer.

Gaidarenko, O. and Xu, Y. (2009). Transcription activity is required for p53-dependent tumor suppression. Oncogene 28: 4397–4401.

Jenkins, L. M., Mazur, S.J., Rossi, M., **Gaidarenko, O.**, Xu, Y., and Appella, E. (2008). Quantitative proteomics analysis of the effects of ionizing radiation in wild type and p53 K317R knock-in mouse thymocytes. Mol Cell Proteomics 7:716-27.

ABSTRACT OF THE DISSERTATION

Carotenoid biosynthesis and productivity in diatoms

by

Olga Gaidarenko

Doctor of Philosophy in Biology

University of California San Diego, 2018

Professor James Golden, Chair

Due to their versatility and modest cultivation requirements, microalgae are a promising potential source of sustainable fuel, chemicals, and food. At present, microalgal production at scale is not economically viable. A major hurdle to productivity is the inefficient use of light energy by dense microalgal cultures. Due to extensive photopigmentation, microalgae closest to the light source absorb more light than they can use, and wastefully dissipate the rest. As a result, light penetrance into the culture is steeply attenuated. Reducing light-harvesting or dissipation capacity of microalgal cells is a promising solution to uneven light distribution in mass

cultures. Most efforts to do so have focused on chlorophytes, with some successes. Diatoms are a class of microalgae that is very promising in terms of productivity and has evolved light-harvesting and photoprotective strategies that differ substantially from those utilized by chlorophytes. This dissertation explores the notion of improving diatom productivity through manipulating their light-harvesting and dissipation capabilities. Because microalgal performance in production conditions can differ substantially from what is observed in the laboratory, the responses of a wild-type production candidate diatom to simulated outdoor conditions are examined in Chapter 1. Substantial diel changes in hypothetical product yields were observed and discussed in terms of what variables need to be optimized to maximize productivity. Main light-harvesting and photoprotective carotenoid-derived photopigments were found to respond differently to chloroplast division and changes in irradiance, suggesting differential regulation. Chapter 2 examined carotenoid biosynthesis in diatoms, because diatom carotenoids play major light-harvesting and photoprotective roles. Targets for genetic manipulation were identified, transgenic lines with two distinct altered photopigmentation phenotypes were generated, and a model for how diatom carotenoid biosynthesis may be differentially regulated in response to chloroplast division and irradiance increase was developed. Chapter 3 focused on examining photosynthetic performance, growth, and productivity of two transgenic strains created in Chapter 2 and identified a strategy that may substantially improve diatom productivity. Overall, the dissertation substantially advances the understanding of diatom carotenoid biosynthesis, identifies strategies for improving light utilization efficiency in diatom cultures, and contributes to the understanding of practices to maximize the productivity of commercial microalgal cultivation.

INTRODUCTION

Photosynthetic microalgae are vastly diverse microorganisms that harness light energy and convert it to biochemical energy. Their unique characteristics with respect to cultivation requirements and chemical composition make them attractive for a variety of commercial applications. Different microalgal species can be adapted to grow in diverse environmental conditions. They can be cultivated on non-arable land and may be able to thrive on waste water and factory flue gases. Microalgae utilize CO_2, light energy, and media with relatively modest nutrient requirements for the biosynthesis of carbohydrates, lipids, proteins, and pigments. Those compounds are exploitable for the purposes of enhancing nutrition for humans and other animals, manufacturing of cosmetics and pharmaceuticals, and providing feedstock for a variety of renewable types of fuel, including biodiesel, hydrogen, methane, and ethanol [Mata et al. 2010, Pultz and Gross 2004, Spolaore et al. 2005].

Given the increasing demand for sustainable and environmentally friendly sources of fuel and food for the ever-growing human population, there have been numerous efforts put forth over several decades in different parts of the globe to domesticate microalgae and take advantage of their modest cultivation requirements and promising production potential. At present, however, the biotechnological efforts remain stymied by the suboptimal performance of microalgae in industrial cultivation conditions [Spolaore et al. 2005]. Microalgal production systems are not yet cost-efficient enough for mass production of microalgal-derived products to be economically feasible [Davis et al. 2011, Stephens et al. 2010]. One of the main factors that limits microalgal productivity in the dense cultures used in commercial settings is the inefficient capture and use of light energy [Goldman 1978, Kok 1960, Neidhardt et al. 1998, Torzillo et al. 2003].

The lack of efficiency with which microalgal cultures utilize light had been noted during early efforts at microalgal biotechnology [Burlew 1953]. Microalgae are only able to utilize approximately 10% of the light energy available to them when exposed to direct sunlight. This

limitation in light utilization is the result of the rate of light absorption at moderate to high photon flux density values greatly surpassing the biochemical conversion rate in microalgal cells [Burlew 1953, De Mooij et al. 2015, Radmer and Kok 1977, Sukenik et al. 1987, Torzillo et al. 2003]. Microalgae assemble extensive pigment-protein complexes known as photoantennae in their thylakoid membranes that harvest light energy and channel it to photosynthetic reaction centers, where it is used to drive photochemistry. Because microalgae are diverse and have complex evolutionary histories, different taxa employ varied strategies to adjust their light-harvesting apparatus in response to light availability. In chlorophytes, for example, the photoantenna size depends on light availability, with larger photoantennae being assembled by cells in more light-limited environments. This increases the surface area available for light absorption and offers a competitive advantage in the wild [Brown and Richardson 1968, Falkowski and Raven 1997, Mitra and Melis 2008, Mussgnug et al. 2007]. Diatoms, brown microalgae of the Stramenopile or heterokont class, do not adjust photoantenna size, but co-regulate the abundance of photoantennae and reaction centers based on cultivation irradiance [Lepetit et al. 2012]. Microalgal cultures that are subject to mixing adjust photopigmentation based on average irradiance. Due to the steep light attenuation in dense cultures **(Fig. I-1)**, the average irradiance is lower than that at the surface, which signals for higher light-harvesting pigmentation than is necessary and useful, wasting cellular resources [Sukenik and Falkowski 1986, Torzillo et al. 2003]. As a result, the extensive light-harvesting pigmentation presents a hurdle to overall productivity of commercially-grown microalgal cultures, which tend to be dense and grown with mixing. Cells that are closest to the light source absorb more light than they are able to utilize. Excess photons are wastefully dissipated as heat and fluorescence, meanwhile light-induced stress results in photoinhibition and further decrease in light utilization efficiency and, therefore, productivity [Myers and Burr 1940, Vonshak and Guy 1992]. Cells that are deeper into the culture are shaded, have less light available

for photosynthesis, and are sub-productive as well. Respiratory losses must also be considered when contemplating the light utilization efficiency in dense cultures. The lower the average light intensity experienced by the cells in the culture is, the greater is the fraction of that light energy that will be necessary to compensate for respiration, leaving a smaller fraction of light energy available for biosynthetic gains [Kok 1953, Kok 1960, Radmer and Kok 1977, Torzillo et al. 2003].

Four general types of solutions to the problem of imbalanced light distribution in dense microalgal cultures have been proposed and attempted since the middle of the 20th century. One of the solutions entailed using high culture density along with rapid mixing, based on the notion that individual cells would be brought to the surface very briefly, then plunged deep into the culture, where there would be practically no light, long enough to assimilate all the harvested light energy prior to being brought to the surface again. Rapidly alternating light/dark periods proved promising in the controlled environment of the laboratory. However, the high variability of outdoor conditions, along with the technical challenges of designing a system that would provide the appropriate light intermittence to a substantial proportion of cells moving randomly in a dense large-scale culture, render the practical applications of the strategy problematic. Increasing cell density has been shown to decrease photoinhibition in outdoor cultures grown with mixing, but not enough to eliminate it in cultures grown at densities that are below the threshold at which productivity becomes compromised due to factors such as autoinhibitor accumulation and respiratory losses [Burlew 1953, Goldman 1978, Kok 1953, Qiang et al. 1996, Richmond 1996, Torzillo et al. 2003]. Another practical consideration is that before the cultures reach the high density necessary for this approach, they would be at lower densities.

Another idea was to use photobioreactor design to dilute the light that impinges directly upon the surface of the culture, and/or distribute it more evenly throughout its depth. Evenari et al. [1953] were among the first to propose such a design, during their early work on microalgal

cultivation in Israel. They were followed by numerous other groups, whose diverse and creative ideas are reviewed in Torzillo et al. [2003]. Whereas, in principle, some improvements in algal productivity appear to be achievable by employing clever photobioreactor design tactics, those are, at present, overshadowed by the prohibitive costs and engineering challenges associated with the construction, scaling up, deployment, and maintenance of such systems [Torzillo et al. 2003].

A system for growing thin-layer dense microalgal cultures in a cascading photobioreactor was developed in the 1960's [Setlik et al. 1970]. The main advantages of that approach are relatively low equipment costs and greater light availability to all the cells in the culture due to the shorter light path, compared to the more commonly used photobioreactors and raceway ponds (6-8 mm versus 15 cm or deeper). Such systems may be promising in terms of productivity. However, factors such as light saturation, photoinhibition, and overabsorption/wasteful dissipation of photons at moderate to high light intensities are not mitigated by the design and present a challenge to optimizing yields. Additionally, care must be taken when designing the pumping mechanism necessary for the circulation of microalgae through that type of system in order to avoid shear stress, which would negatively impact productivity [Doucha and Livansky 2009, Torzillo et al. 2010]. Again, the practicality of growing the cultures from lower density inoculates to the desired high density must be considered with this approach.

Finally, the reduction of cellular light-harvesting capacity had been suggested as a means to create microalgal strains with improved light utilization efficiency at high culture densities [Radmer and Kok 1977, Sukenik and Falkowski 1987]. In theory, cells with reduced light-harvesting capacity would become light-saturated and risk photoinhibition at greater light intensities than their wild-type (WT) counterparts. In a dense culture of cells with reduced light-harvesting capacity, those closest to the light source would absorb less light energy and process it with higher efficiency. Light penetrance into the culture would increase, allowing for a greater proportion of the cells to be

photosynthetically active. Thus, the culture as a whole would be more productive and waste less light energy through dissipation due to overabsorption. Also, with light availability being less of a limiting factor, mass cultures may be able to achieve higher densities for a given culture depth [De Mooij et al. 2015, Mitra and Melis 2008].

This principle was first experimentally demonstrated by Melis et al. [1999] in the late 1990's. They used high light-adapted *Dunaliella salina* (chlorophyte) as a model system to study the effects of having reduced photoantenna size on a culture as a whole. Compared to low light-adapted cells with substantially more extensive photoantennae, the high light-adapted cells indeed allowed for greater light penetrance into the culture, required much higher irradiance for light-saturation, and had a higher per chlorophyll photosynthetic efficiency. High light-acclimated cells would not be able to sustain their small photoantenna phenotype at higher culture densities, and thus would not be an option for improving productivity in mass cultures. However, these findings encourage genetic manipulation of microalgae as an approach to creating strains with a stable reduction in light-harvesting capacity with the hopes that they will prove to be more productive in mass cultures [Melis et al. 1999, Neidhardt et al. 1998].

Around the same time frame, Nakajima and colleagues published the first studies of a photoantenna mutant. They demonstrated that a *Synechocystis* sp. (cyanobacteria) mutant with diminished photoantennae was less susceptible to photoinhibition, became light-saturated at higher irradiance, and was more photosynthetically active and productive than WT under their cultivation conditions [Nakajima and Ueda 1997, Nakajima et al. 1999, Nakajima and Ueda 1999]. Shortly after, more studies by various groups followed, characterizing various chlorophyte and cyanobacterial strains with the reduced photoantenna phenotype. Additionally, one study featured a photopigment mutant of a *Cyclotella* sp. (diatom) [Huesemann et al. 2009]. The cultivation conditions and parameters assessed varied greatly between the different studies. Cultivation

systems ranged from small laboratory photobioreactors to 7 L hanging-bag photobioreactors grown outdoors. One or more improvements over parental strains, such as greater light penetrance into the culture, light saturation at higher irradiance, increased maximum photosynthetic rate, increased maximum cell density, improved light to biomass conversion efficiency, and decrease in photoinhibition and photon overabsorption were noted for most of the strains [Beckmann et al. 2009, Cazzaniga et al. 2014, Kirst et al. 2012, Kwon et al. 2013, Lea-Smith et al. 2014, Leganes et al. 2014, Mitra and Melis 2008, Mussgnug et al. 2007, Nakajima et al. 1999, Nakajima and Ueda 1997, Nakajima and Ueda 2000, Ort et al. 2011, Polle et al. 2000, Polle et al. 2001, Polle et al. 2002, Polle et al. 2003]. Some groups saw increased growth rates [Beckmann et al. 2009, Mussgnug et al. 2007], while others reported slower growth [Huesemann et al. 2009, Kirst et al. 2012]. Some groups measured higher productivity under their cultivation conditions [Cazzaniga et al. 2014, Nakajima and Ueda 1999, Nakajima and Ueda 2000, Nakajima et al. 2001, Polle et al. 2000, Polle et al. 2003]. One study found that a truncated photoantenna mutant of the chlorophyte *Chlamydomonas reinhardtii* produced hydrogen more efficiently than WT in their experimental setup [Kosorouv et al. 2011]. Others found that the improved photosynthetic characteristics of their lines did not translate to greater productivity in their culture conditions [Huesemann et al. 2009, Lea-Smith et al. 2014, Page et al. 2012]. It is likely that for strains generated through chemical or ultraviolet radiation-induced mutagenesis, multiple other genetic loci were affected, which may have had an adverse impact on the fitness and performance of those strains. Insertional or transposon-mediated mutagenesis, as well as the introduction of a transgene, are substantially safer in terms of the chances of introducing unwanted genetic alterations along with the intended one, since there usually is only one or few insertion sites. Performance of the strains generated using those approaches may have been also affected by unintended consequences of the genetic alterations

that led to the desired phenotypic changes [Cazzaniga et al. 2014, de Mooij et al. 2015, Huesmann et al. 2009, Mitra and Melis 2008].

De Mooij et al. [2015] recently tested the performance of four *C. reinhardtii* photoantenna mutants in simulated mass culture conditions, using a laboratory scale photobioreactor and irradiance of 1500 µmol photons $m^{-2} s^{-1}$, approximating that of full sunlight. Despite the promising initial characterization of the mutants [Bonente et al. 2011, Kirst et al. 2012, Mussgnug et al. 2007], higher productivity was not observed in their culture system. Compared to WT, a greater sensitivity to high light and tendency towards photodamage were observed in the four mutant lines. This finding could be attributed to off-target effects of the procedures used to obtain the mutants, a loss of light-responsive regulatory capabilities in the mutant photoantenna complexes [de Mooij et al. 2015], or a general reduction in antioxidant capacity due the decrease in pigmentation, with a concomitant increase in susceptibility to damage caused by reactive oxygen species that are produced in the thylakoid membranes when exposed to high levels of irradiance [Baroli et al. 2004, Li et al. 2012, Torzillo et al. 2003, Vetoshkina et al. 2015].

It is important to note that whereas studies in laboratory conditions are informative, they do not necessarily predict outcomes in mass culture conditions, which can differ substantially with respect to the quality and quantity of available light, light path/culture depth, pond/photobioreactor geometry and its effects on cell mixing, and diurnal light/temperature fluctuations experienced by outdoor cultures. To date, few lines with altered light-harvesting capabilities have been examined in production conditions. The mutant *Cyclotella* sp. (diatom) generated by Huesemann et al. [2009] performed worse than the WT in a raceway pond. However, it was generated by two rounds of mutagenesis, once with ethylmethylsulfonate, then with ultraviolet radiation, and likely harbored numerous other mutations that gave it an overall reduced fitness, as indicated by its slower growth rate and higher propensity for washout in semi-continuous

culture conditions. The mutant was also green rather than golden-brown, suggesting that a substantial loss of carotenoids, critical for photoprotection and reactive oxygen species scavenging, had occurred [Huesmann et al. 2009]. On the other hand, Cazzaniga et al. [2014] reported greater biomass productivity in their mutant chlorophyte *Chlorella sorokiniana* in laboratory conditions as well as outdoors when cultivated in 7 L hanging bag photobioreactors.

Improving productivity in mass cultures by altering cellular light-harvesting properties remains an attractive prospect. The challenges discovered to date are informative and must be taken into account for future work. It will be important to develop methods for reducing the cellular light-harvesting capacity that do not adversely affect overall cellular fitness. Tolerance of high and fluctuating light conditions should also be prioritized [De Mooij et al. 2015]. In addition to modifying light harvesting, reducing light dissipation may be a valuable strategy for improving light utilization efficiency in mass cultures. Berteotti et al. [2016] recently found that downregulating light energy dissipation through non-photochemical quenching (NPQ) in *C. reinhardtii* improved biomass productivity in a small scale photobioreactor.

Parental strain characteristics must be considered when selecting species for further development of these ideas. Cyanobacteria and eukaryotic microalgae have unique advantages and should be developed in parallel. The former are propitious for the production of small molecules that can be secreted, such as ethanol and fatty acids. The latter have desirable attributes for the production of compounds that require more storage space, such as lipids, proteins, and carbohydrates, as well as the expression of heterologous proteins that require eukaryotic post-translational modifications [Wijffels et al. 2013]. Diatoms, which have so far been understudied for the problem of inefficient light utilization in mass cultures, are an attractive class of eukaryotic microalgae for selecting candidates. Due to their distinctive metabolism, high productivity, and environmental success, numerous species of diatoms are expected to be valuable for commercial

applications such as biofuels [Hildebrand et al. 2012, Sheehan et al. 1998]. Based on available studies, transgenes in diatoms appear to be more stable than in chlorophytes [Hildebrand et al. 2012, Kumar 2015]. Diatom photopigment composition, organization of thylakoid membranes and light harvesting complexes, as well as strategies for short- and long-term adaptation to changes in irradiance differ from those in chlorophytes [Wilhelm et al. 2006]. Due to the differences in light-harvesting and photoprotective mechanisms, diatoms are more efficient at dissipating excess energy and more productive than chlorophytes in fluctuating light conditions, such as those that would be experienced in mass cultures [Hildebrand et al. 2012, Wagner et al. 2006]. Additionally, diatoms are more efficient than chlorophytes at absorbing blue-green light, which is important in aquatic environments. This difference is due to the fact that the main accessory photopigment in diatoms enables absorption between 460 and 570nm, a range that is absent in chlorophytes.

The overarching objective of this thesis was to contribute to the understanding of how diatom productivity can be improved. Because the ultimate goal of strain improvement is to increase productivity in production conditions, Chapter 1 explored diel physiological and metabolic changes in the production candidate diatom *Cyclotella cryptica* (WT) under simulated outdoor conditions. A benchtop photobioreactor **(Fig. I-1)** was used to subject it to sinusoidal diel changes in light and temperature. Synchronous cell cycle progression, typical of microalgal cultures grown on a light-dark regime, was observed. Diel changes in optical density (a proxy for biomass), neutral lipid triacylglycerol (TAG, of interest for biofuel production), and photopigments were recorded and discussed in relation to cell cycle progression as well as diel changes in light and temperature. Time of day substantially affected potential product yields. The necessity of taking specific cultivation conditions into consideration when evaluating strain performance, variables that must be optimized to enhance product yields, and differences that may be expected based on microalgal taxa were discussed. Additionally, cellular photopigment abundance was assessed at high

resolution, and the findings suggested that the diatom carotenoid biosynthesis pathway may be differentially regulated in response to chloroplast replication and changes in irradiance.

Chapter 2 focused on elucidating the carotenoid biosynthesis pathway in the model diatom *Thalassiosira pseudonana*. In diatoms, carotenoids play a major role in light harvesting and dissipation. Thus, it is promising to explore manipulating diatom carotenoid biosynthesis in efforts to improve light utilization efficiency in diatom cultures. The identities of multiple enzymes that catalyze key biosynthetic steps, as well as the sequence of final steps leading to the main diatom light-harvesting and photoprotective carotenoids, have not been identified. Therefore, the goal of Chapter 2 was to improve the understanding of diatom carotenoid biosynthesis and identify targets for genetic manipulation. Basic bioinformatic analyses combined with available transcriptomic and physiological data were employed. A major finding of Chapter 2 was the identification of a novel violaxanthin de-epoxidase-like enzyme (VDL2), found to catalyze the biosynthesis of the main diatom accessory light-harvesting pigment fucoxanthin (Fx). Overexpressing VDL2 resulted in an increase of cellular Fx content accompanied by a stoichiometric decrease in the main photoprotective pigment pool, diadinoxanthin (Ddx) and its reversibly de-epoxidized form diatoxanthin (Dtx), which is necessary for the majority of light energy dissipation. Another major finding was that reducing Fx results in a coordinate reduction of all photosynthetic pigments, including chlorophylls, while photopigment ratios are preserved. A model for how the diatom carotenoid biosynthesis pathway may be differentially regulated in response to chloroplast division and irradiance increase was developed.

Chapter 3 examined photosynthetic parameters, growth, carbon partitioning, and productivity in two transgenic *T. pseudonana* lines, one overexpressing VDL2, and the other with a total reduction of photosynthetic pigmentation, as described above. A major finding of this chapter was that reducing photoprotective pigments (Ddx+Dtx) is a promising strategy for increasing diatom

productivity. *T. pseudonana* overexpressing VDL2 accumulated up to 3.4 times as much TAG and twice as much protein as WT during exponential growth. The accumulation of TAG and protein inversely correlated with NPQ, which is mainly a function of Ddx+Dtx content. VDL2 overexpression lines grew up to 7% slower than WT, but harvesting could be timed so as to take advantage of their higher yield potential. Possible reasons for the observed slower growth and potential strategies to improve upon the phenotype were discussed. A more modest increase in TAG accumulation of up to 40% more than in WT was observed in *T. pseudonana* with an overall reduction of photosynthetic pigments, obtained by using antisense to simultaneously silence both copies of the LUT1-like gene. The cells were 5-10% smaller than WT, contained 11-19% less protein, and appeared stressed, possibly due to a cell cycle progression defect, but growth rate in our cultivation conditions was comparable to WT.

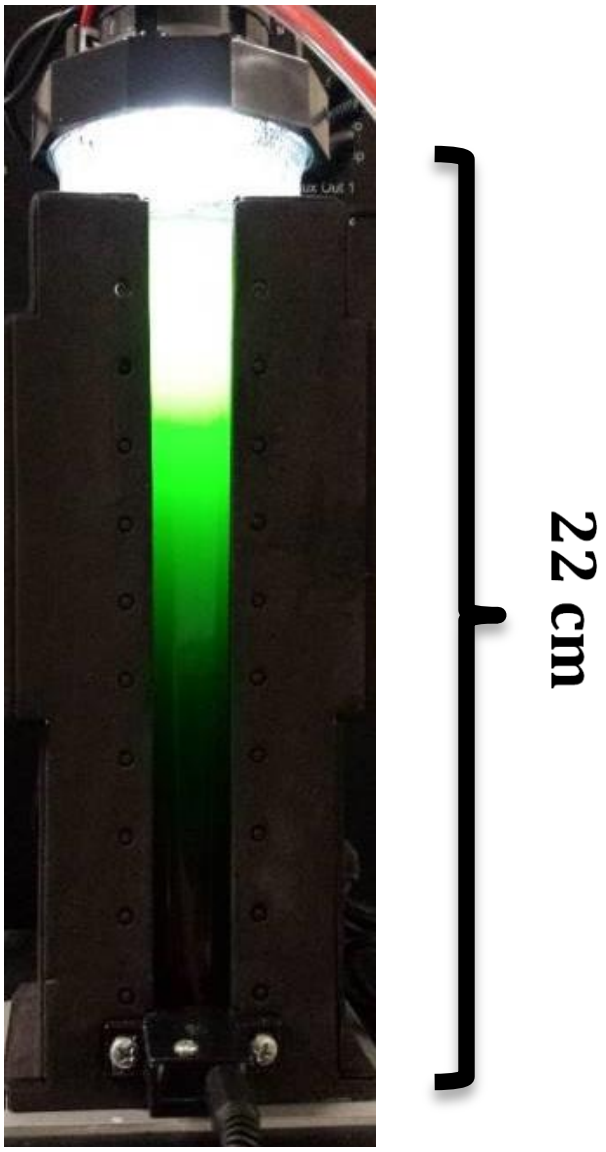


Figure. I.1.
22 cm-deep culture of *Chlorella* sp. in a bench-top photobioreactor. Despite the incident light intensity approximating full sunlight, cells at the bottom of the culture are in darkness.

REFERENCES FOR THE INTRODUCTION

Baroli I., Gutman B. L., Ledford H. K., Shin J. W., Chin B. L., Havaux M., Niyogi K. K. (2004) Photo-oxidative stress in a xanthophyll-deficient mutant of *Chlamydomonas*. J Biol Chem 279(8):6337-6344.

Beckmann J., Lehr F., Finazzi G., Hankamer B., Posten C., Wobbe L., Kruse O. (2009) Improvement of light to biomass conversion by de-regulation of light-harvesting protein translation in *Chlamydomonas reinhardtii*. J Biotechnol 142(1):170-177.

Berteotti S., Ballottari M., Bassi R. (2016) Increased biomass productivity in green algae by tuning non-photochemical quenching. Sci Rep 6:21339.

Bonente G., Formighieri C., Mantelli M., Catalanotti C., Giuliano G., Morosinotto T., Bassi R. (2011) Mutagenesis and phenotype selection as a strategy toward domestication of *Chlamydomonas reinhardtii* strains for improved performance in photobioreactors. Photosynth Res 108: 107-120.

Brown T. E., Richardson F. L. (1968) The effect of growth environment on the physiology of algae: light intensity. J Phycol 4:38-54.

Burlew J. S. (1953) Current status of the large-scale cultivation of algae. pp 3-23. In: Burlew JS(ed). *Algal Culture: From Laboratory to Pilot Plant*. Carnegie Institution of Washington Publication 600, USA.

Cazzaniga S., Dall'Osto L., Szaub J., Scibilia L., Ballottari M., Purton S., Bassi R. (2014) Domestication of the green alga *Chlorella sorokiniana*: reduction of antenna size improves light-use efficiency in a photobioreactor. Biotechnol Biofuels 7:157.

Davis R., Aden A., Pienkos P. T. (2011) Techno-economic analysis of autotrophic microalgae for fuel production. Appl Energ 88:3524-3531.

De Mooij T., Janssen M., Cerezo-Chinarro O., Mussgnug J. H., Kruse O., Ballottari M., Bassi R., Bujaldon S., Wollman F. A., Wijffels R. H. (2015) Antenna size reduction as a strategy to increase biomass productivity: a great potential not yet realized. J Appl Phycol 27(3):1063-1077.

Doucha J., Livansky K. (2009) Outdoor open thin-layer microalgal photobioreactor: potential productivity. J Appl Phycol 21:111-117.

Evenari M., Mayer A. M., Gottesman E. (1953) Experiments on culture of algae in Israel. pp197-203. In: Burlew JS(ed). *Algal Culture: From Laboratory to Pilot Plant*. Carnegie Institution of Washington Publication 600, USA.

Falkowski P. G., Raven J. A. (1997) The molecular structure of the photosynthetic apparatus. pp 163-192. In: *Aquatic Photosynthesis*. Capital City Press, USA.

Goldman, J. C. (1978) Outdoor algal mass cultures – II. Photosynthetic yield limitations. Water Res 13:119-136.

Hildebrand M., Davis A. K., Smith S. R., Traller J. C., Abbriano R. (2012) The place of diatoms in the biofuels industry. Biofuels 3(2):221-240.

Huesemann M. H., Hausmann T. S., Bartha R., Aksoy M., Weissman J. C., Benemann J. R. (2009) Biomass productivities in wild type and pigment mutant of *Cyclotella* sp. (diatom). Appl Biochem Biotechnol 157:507-526.

Kirst H., Garcia-Cerdan J. G., Zurbriggen A., Melis A. (2012) Assembly of the light-harvesting chlorophyll antenna in the green alga *Chlamydomonas reinhardtii* requires expression of the TLA2-CpFTSY gene. Plant Physiol 158(2):930-945.

Kok B. (1953) Experiments of photosynthesis by *Chlorella* in flashing light. pp 63-75. In: Burlew JS(ed). *Algal Culture: From Laboratory to Pilot Plant*. Carnegie Institution of Washington Publication 600, USA.

Kok B. (1960) Efficiency of photosynthesis. pp 566-633. In: Pirson A(ed). *Encyclopedia of Plant Physiology, Vol 5: The Assimilation of CO_2.* Springer-Verlag Berlin Heidelberg, Germany.

Kosourov S. N., Ghirardi M. L., Seibert M. (2011) A truncated antenna mutant of *Chlamydomonas reinhardtii* can produce more hydrogen than the parental strain. Int J Hydrogen Energ 36(3):2044-2048.

Kumar S. (2015) GM algae for biofuel production: biosafety and risk assessment. Coll Biosaf Rev 9:52-75.

Kwon J. H., Bernat G., Wagner H., Rogner M., Rexroth S. (2013) Reduced light-harvesting antenna: consequences on cyanobacterial metabolism and photosynthetic productivity. Algal Res 2(3):188-195.

Lea-Smith D. J., Bombelli P., Dennis J. S., Scott S. A., Smith A. G., Howe C. J. (2014) Phycobilisome-deficient strains of *Synechocystis* sp. PCC 6803 have reduced size and require carbon-limiting conditions to exhibit enhanced productivity. Plant Physiol 165(2):705-714.

Leganes F., Martinez-Granero F., Munoz-Martin M. A., Marco E., Jorge A., Carvajal L., Vida T., Gonzalez-Pleiter M., Fernandez-Pinas F. (2014) Characterization and responses to environmental cues of a photosynthetic antenna-deficient mutant of the filamentous cyanobacterium *Anabaena* sp. PCC 7120. J Plant Physiol 171:915-926.

Lepetit B., Goss R., Jakob T., Wilhelm C. (2012) Molecular dynamics of the diatom thylakoid membrane under different light conditions. Photosynth Res 111(1-2): 245-257.

Li Z., Kiesling J. D., Niyogi K. K. (2012) Overlapping photoprotective function of vitamin E and carotenoids in *Chlamydomonas*. Plant Physiol 158(1):313-323.

Mata T. M., Martins A. A., Caetano N. S. (2010) Microalgae for biodiesel production and other applications: a review. Renew Sust Energ Rev 14: 217-232.

Melis A., Neidhardt J., Benemann J. R. (1999) *Dunaliella salina* (Chlorophyta) with small chlorophyll antenna sizes exhibit higher photosynthetic productivities and photon use efficiencies than normally pigmented cells. J Appl Phycol 10:515-525.

Mitra M., Melis A. (2008) Optical properties of microalgae for enhanced biofuels production. Opt Express 16(26):21807-21820.

Mussgnug J. H., Thomas-Hall S., Rupprecht J., Foo A., Klassen V., McDowall A., Schenk P. M., Kruse O., Hankamer B. (2007) Engineering photosynthetic light capture: impacts on improved solar energy to biomass conversion. Plant Biotech J 5:802-814.

Myers J., Burr G. O. (1940) Studies on photosynthesis: some effects of light of high intensity on *Chlorella*. J Gen Physiol 24:45-67.

Nakajima Y., Tsuzuki M., Ueda R. (1999) Reduced photoinhibition of a phycocyanin-deficient mutant of *Synechocystis* PCC 6714. J Appl Phycol 10:447-452.

Nakajima Y., Ueda R. (1997) Improvement of photosynthesis in dense microalgal suspension by reduction of light harvesting pigments. J Appl Phycol 9:503-510.

Nakajima Y., Ueda R. (1999) Improvement of microalgal photosynthetic productivity by reducing the content of light harvesting pigment. J Appl Phycol 11:195-201.

Nakajima Y., Ueda R. (2000) The effect of reducing light-harvesting pigment on marine microalgal productivity. J Appl Phycol 12:285-290.

Neidhardt J., Benemann J. R., Zhang L., Melis A. (1998) Photosystem-II repair and chloroplast recovery from irradiance stress: relationship between chronic photoinhibition, light-harvesting chlorophyll antenna size and photosynthetic productivity in *Dunaliella salina* (green algae). Photosynth Res 56:175-184.

Ort D. R., Zhu X., Melis A. (2011) Optimizing antenna size to maximize photosynthetic efficiency. Plant Physiol 155:79-85.

Page L. E., Liberton M., Pakrasi H. B. (2012) Reduction of photoautotrophic productivity in the cyanobacterium *Synechocystis* sp. strain PCC 6803 by phycobilisome antenna truncation. Appl Environ Microbiol 78(17):6349-6351.

Polle J. E. W., Benemann J. R., Tanaka A., Melis A. (2000) Photosynthetic apparatus organization and function in the wild type and a chlorophyll *b*-less mutant of *Chlamydomonas reinhardtii*. Dependence on carbon source. Planta 211:335-344.

Polle J. E. W., Kanakagiri S., Jin E. S., Masuda T., Melis A. (2002) Truncated chlorophyll antenna size of the photosystems – a practical method to improve microalgal productivity and hydrogen production in mass culture. Int J Hydrogen Energ 27:1257-1264.

Polle J. E. W., Kanakagiri S. D., Melis A. (2003) *tla1*, a DNA insertional transformant of the green alga *Chlamydomonas reinhardtii* with a truncated light-harvesting chlorophyll antenna size. Planta 217:49-59.

Polle J. E. W., Niyogi K. K., Melis A. (2001) Absence of the pigments lutein, violaxanthin and neoxanthin affects the functional chlorophyll antenna size of the photosystem-II but not that of photosystem-I in the green alga *Chlamydomonas reinhardtii*. Plant Cell Physiol 42(5):482-491.

Pultz O., Gross W. (2004) Valuable products from biotechnology of microalgae. Appl Microbiol Biotechnol 65:635-648.

Qiang H., Guterman H., Richmond A. (1996) Physiological characteristics of *Spirulina platensis* (cyanobacteria) cultured at ultrahigh cell densities. J Phycol 32:1066-1073.

Radmer R., Kok B. (1977) Photosynthesis: Limited yields, unlimited dreams. BioScience 27(9):599-605.

Richmond A. (1996) Efficient utilization of high irradiance for production of photoautotrophic cell mass: a survey. J Appl Phycol 8:381-387.

Setlik I., Veladimir S., Malek I. (1970) Dual purpose open circulation units for large scale culture of algae in temperate zones. I. Basic design considerations and scheme of pilot plant. Algol Stud 1: 111-164.

Sheehan J., Dunahay T., Benemann J., Roessler P. (1998) *A look back at the U.S. Department of Energy's Aquatic Species Program: biodiesel from algae; close-out report.* United States. Doi:10.2172/15003040.

Spolaore P., Joannis-Cassan C., Duran E., Isambert A. (2005) Commercial applications of algae. J Biosci Bioeng 101(2):87-86.

Stephens E., Ross I. L., King Z., Mussgnug J. H., Kruse O., Posten C., Borowitzka M. A., Hankamer B. (2010) An economic and technical evaluation of microalgal biofuels. Nat Biotechnol 28:126-128.

Sukenik A., Bennett J., Falkowski P. (1987) Light-saturated photosynthesis – limitation by electron transport or carbon fixation? Biochim Biophys Acta 891:205-215.

Sukenik A., Falkowski P. G. (1986) Potential enhancement of photosynthetic energy conversion in algal mass culture. Biotechnol Bioeng 30:970-977.

Torzillo G., Giannelli L., Martinez-Roldan A. J., Verdone N., De Filippis P., Scarsella M., Bravi M. (2010) Microalgae culturing in thin-layer photobioreactors. Chem Eng Transact 20:265-270.

Torzillo G., Pushparaj B., Masojidek J., Vonshak A. (2003) Biological constraints in algal biotechnology. Biotechnol Bioprocess Eng 8:338-348.

Vetoshkina D. V., Borisova-Mubarakshina M. M., Naydov I. A., Kozuleva M. A., Ivanov B. N. (2015) Impact of high light on reactive oxygen species production within photosynthetic biological membranes. J Biol Life Sci 6(2):50-60.

Vonshak A., Guy R. (1992) Photoadaptation, photoinhibition, and productivity in the blue-green alga, *Spirulina platensis* grown outdoors. Plant Cell Environ 15:613-616.

Wagner H., Jacob T., Wilhelm C. (2006) Balancing the energy flow from captured light to biomass under fluctuating light conditions. New Phytol 169:95-108.

Wijffels R. H., Kruse O., Hellingwerf K. J. (2013) Potential of industrial biotechnology with cyanobacteria and eukaryotic microalgae. Curr Opin Biotechnol 24:405-413.

Wilhelm C., Buchel C., Fisahn J., Goss R., Jakob T., LaRoche J., Lavaud J., Lohr M., Riebesell U., Stehfest K., Valentin K., Kroth P. G. (2006) The regulation and nutrient assimilation in diatoms is significantly different from green algae. Protist 157(2): 91-124.

CHAPTER 1

Timing is everything: diel metabolic and physiological changes in the diatom *Cyclotella cryptica* grown in simulated outdoor conditions

1.1 ABSTRACT

Microalgal cultures grown on a light-dark cycle experience diel patterns in metabolic and physiological processes, including cell cycle synchronization, but the implications for productivity in terms of biomass and commercially-appealing molecules are not commonly appreciated. Despite a long history of diel response studies, only recently have photobioreactor technology advances enabled the use of sinusoidal light and temperature to more accurately mimic outdoor conditions. The present study investigates cell cycle progression and dynamic changes in optical density as a proxy for biomass, triacylglycerol (TAG), and photosynthetic pigments on a 24-hour scale in the diatom *Cyclotella cryptica* grown using a sinusoidal light and temperature regime. Cell division synchronized to occur predominantly in the middle of the light period and biomass started to increase several hours earlier, as the cells prepared to divide. TAG levels increased during the day and decreased at night, with a mid-day dip corresponding to the time when lipid needs for cell division-associated membrane biosynthesis would be high. For the first time, photosynthetic pigment dynamics data, obtained with higher temporal resolution than previously reported for microalgae, was overlaid with cell cycle progression, indicating that while some photosynthetic pigments respond primarily to light, others are influenced by the cell cycle. Additionally, our results indicate that in a synchronized culture, potential product yields change substantially throughout the day. This may inform harvest timing to significantly increase yield.

1.2 INTRODUCTION

Microalgae are a large and very diverse group of unicellular photosynthetic organisms that occupy a wide variety of environmental niches and have major ecological significance [Ebenzer et al. 2012, Sumi 2009]. They are also of interest commercially and can be used to obtain biomass and

a variety of products such as lipids, proteins, carbohydrates, pigments, and other functional nutrients, which can be used for diverse applications such as biofuels, food supplementation, cosmetics, and bioplastics [Khanra et al. 2018, Matos et al. 2017]. With the exception of polar dwellers, natural populations of microalgae typically experience regular and predictable light-dark cycles and concomitant changes in temperature. In response, individual cells exhibit cyclical patterns in metabolic and physiological processes and there is population-level synchronization, allowing for efficient use of resources [Falkowski and Raven 2007]. A major event in a cell's life that influences the timing of many of those changes is division, and most microalgal species synchronize their cell cycles when exposed to light-dark regimes. Several microalgal lineages, including green and red algae, appear to have the cell cycle gated by an endogenous circadian mechanism, restricting division to certain times of day (typically night) [Miyagishima 2017]. This has been found to not generally be the case with diatoms [Chisholm and Brand 1981], which may divide at different times of day and in some cases exhibit no periodicity in division, depending on the species and the cultivation conditions [Chisholm et al. 1980, Nelson and Brand 1979, Paasche 1968, Williamson 1980]. The cyclic nature of the various processes has important implications for work with microalgae, from the field, where timing of measurements and sample collection must be taken into consideration, to the laboratory, especially when evaluating and/or developing strains for outdoor production, to production systems, where the optimal timing of dilution, nutrient addition, and harvest should be determined. There have been numerous informative studies examining growth, productivity, and metabolism of microalgal cultures entrained to various light-dark regimes in the laboratory, but the majority of them have been carried out in constant temperature and using a square-wave (on/off) light regime [Anderson and Sweeney 1977, Chauton et al. 2013, Chisholm et al. 1980, Clark et al. 2002, Cuhel et al. 1984, Darley and Volcani 1971, Eppley et al. 1967, Hoogenhout 1963, Humphrey 1979, Joseph and Villareal 1998, Kohata and Watanabe 1988,

Kohata and Watanabe 1989, Nelson and Brand 1979, Owens et al. 1980, Paasche 1968, Post et al. 1984, Ragni and d'Alcala 2007, Raimbault and Mingazzini 1987, Rivkin 1985, Sukenik and Carmeli 1990, Varum et al. 1986, Williamson 1980]. Whereas those studies greatly advanced the general understanding of the effects of light-dark cycling on microalgae, they did not accurately mimic the sinusoidal light and temperature conditions microalgae experience in the wild and in outdoor cultivation systems. Orefice et al. [2016], utilizing the diatom *Skeletonema marinoi*, demonstrated that a square-wave light regime results in significant differences in cell physiology and metabolism when compared to a sinusoidal light regime. Additionally, a difference of 10°C affects an approximately twofold change in metabolic reaction rates in microalgae [Raven and Geider 1988], and it has been demonstrated that diel temperature variation has an impact on microalgal growth and productivity [Bonnefond et al. 2016, Edmundson and Huesemann 2015, Ogbonna and Tanaka 1996, Yang et al. 2016]. Recently, several groups have taken advantage of technological advances and examined various responses of different microalgal species to sinusoidal light and in some cases temperature, beginning to forge a better understanding of what happens outdoors, but under more controlled conditions [Bonnefond et al. 2016, Jallet et al. 2016, Lacour et al. 2012, Tamburic et al. 2014]. The present study is the first to apply sinusoidal temperature changes along with a sinusoidal light regime to a diatom, examining diel metabolic and physiological changes in *Cyclotella cryptica* in a bench-top photobioreactor. Triacylglycerol (TAG) abundance, optical density as a proxy for biomass, and photosynthetic pigments were tracked over a 24-hour period and related to cell cycle progression. This allowed a distinction to be made between the effects of the cultivation regime and cell cycle progression on some of the parameters, such as TAG content and the abundance of different photosynthetic pigments, and identified different optimal times for harvesting to maximize productivity.

1.3 MATERIALS AND METHODS

1.3.1 Cultivation Conditions

C. cryptica CCMP332, collected in 1956 from West Tisbury Great Pond, Martha's Vineyard, Massachusetts USA, was used in this study. The lab stock culture was maintained in 125 mL Erlenmeyer flasks at 18°C under continuous cool-white fluorescent illumination of 150 µmol $m^{-2}sec^{-1}$ in artificial sea water (ASW) medium [Darley and Volcani 1969]. The stock was inoculated into 750 mL ASW at 1.0 x 10^5 cells/mL for Experiment 1 (Exp. 1) and 1.5 x 10^4 cells/mL for Experiment 2 (Exp. 2), for cultivation in the Phenometrics environmental photobioreactor (ePBR) 1.1 [Lucker et al. 2014] with continuous stirring set to 400 rpm and gently bubbled air. The ePBR was programmed for a 12h:12h light-dark cycle, with a sinusoidal light regime maximizing at 2000 µmol photons m^{-2} sec^{-1}, approximating full sunlight, and a sinusoidal temperature regime spanning 14-28°C, a range comfortable for *C. cryptica* **(Fig. 1.1)**. The offset between changes in light and temperature was determined automatically by the software.

For equilibration to the growth conditions prior to intensive sampling and to ensure maintenance of synchronization [Hoogenhout 1963], the cultures were diluted nightly based on cell counts to maintain a slightly past mid-exponential phase once the inoculum had reached the desired cell concentration **(Fig. S1.1)**. The appropriate volume of culture was removed and replaced with fresh ASW through an opening in the lid using a sterilized cannula. The concentration was chosen to be high enough to allow for efficient sampling while maintaining exponential growth of the culture, and mimics the dense cultures used in production systems. For Exp. 1, the total pre-sampling equilibration period was 9 days, with dilutions performed for the last 7 days, 4.3±0.8 hours into the dark period. For Exp. 2, the total pre-sampling equilibration period was 11 days, with dilutions performed 2.9±0.2 hours into the dark period for the last 7 days. After the equilibration

period, the cultures were intensively sampled for 24 hours without dilution, starting 1 hour prior to dawn.

1.3.2 Cell Concentration and Curve Fitting

Cell counts were performed with the MUSE® Cell Analyzer (EMD Millipore, Billerica, MA) and spot-checked with a hemocytometer. For the intensive sampling periods, curve fitting was performed using online software available from MyAssays Ltd. at https://www.mycurvefit.com. Four-parameter logistic fit was used, except for the last 3 and 2 points of Exp. 1 and Exp. 2, respectively. Those points reflect cell concentration increase due to a subpopulation of cells that divided at night, and were fitted with a straight line **(Fig. S1.2)**.

1.3.3 OD750 Measurements

Optical density at 750 nm (OD750) was automatically and continuously measured at approximately 10 min intervals by the ePBR. Because the intervals were not exactly 10 min, the timing of measurements varied slightly from day to day. For dawn and dusk analyses (1.4.3.4 and 1.4.3.5), values for exactly 10 min intervals were interpolated from the raw data using Origin software (OriginLab, Northampton, MA).

1.3.4 OD750 as a Proxy for Biomass

Cultivation experiments with *C. cryptica* in climate simulation raceway ponds at the Pacific Northwest National Laboratory [Huesemann et al. 2017, 2018], indicated a strong correlation

between OD750 and ash-free dry weight (AFDW), a measure of biomass **(Fig. S1.3)**. Duplicate samples were taken from replicate ponds at least twice a day over a 6-day period where the cultures ranged from lag phase, through exponential, to the beginning of stationary phase. Some of the samples were taken at night, and intensive sampling days during which cell division synchronization was observed were included in the analysis. The robust linear relationship was observed despite the presumable differences in cellular characteristics at different sampling points, suggesting that for actively growing *C. cryptica*, OD750 is a reliable proxy for AFDW.

1.3.5 Cell Cycle Analysis

One mL of culture was taken for cell cycle analysis at various times throughout the intensive sampling days and processed as previously described [Abbriano et al. 2018].

1.3.6 Cellular TAG Content Analysis

One mL of culture was taken for relative cellular TAG content analysis via the fluorescent dye 4,4-difluoro-4-bora-3a,4a-diaza-s-indacene (BODIPY 493/503, Molecular Probes) staining at various times throughout the intensive sampling days and processed as previously described [Traller and Hildebrand 2013]. The approach was chosen over gravimetric methods because doing *in situ* analysis allows direct measurement of cellular lipid content in real time, which eliminates errors due to extraction non-specificity, measurement of small weights, and losses during multi-step procedures, and is thus more accurate [Bono et al. 2015]. Additionally, the ePBR does not hold enough culture volume to sample for analysis by gravimetry.

1.3.7 Photosynthetic Pigment Analysis

The cultures were sampled for photosynthetic pigments at various times throughout the intensive sampling days, with more frequent sampling around dawn and dusk. Between 1.5-3x10^6 cells per timepoint were harvested by filtration though a GF/F filter and immediately frozen and stored in liquid nitrogen. Pigments were extracted and analyzed by high-performance liquid chromatography (HPLC) as previously described [Kozlowski et al. 2011]. Cellular pigment content was calculated by normalizing HPLC-derived concentrations by the number of cells harvested using the fitted curves described in 1.3.2.

1.3.8 Statistics

Our study employed two replicate cultures. OD750 measurements were performed automatically approximately every 10 min, resulting in continuous curves that did not present outlier data points. For cell cycle and cellular TAG content analyses, 10^4 cells were interrogated. Due to typical high cell-to-cell variation in TAG (e.g., Traller et al. 2013), using standard deviation (SD) to evaluate error is not useful, whereas standard error gives an accurate measure of the mean. Given the large number of cells evaluated, standard error is extremely small. Cell counts performed with the MUSE® Cell Analyzer used the average of 1x10^3 cells in triplicate to determine concentration. Spot-checking with a hemocytometer ensured accuracy. As described in 1.3.2, curve fitting was used to smooth the curves and ensure consistent values for the analyses. Due to sampling frequency limitations, only one pigment sample per time-point was obtained; however, as detailed in 1.5.5, differences in the patterns of change between different pigments provide an internal control, demonstrating that the observed dynamics were not due to sampling or processing error. Additionally, to assess error that may occur as a result of sampling and processing, we performed a

control experiment in which a culture of *C. cryptica* was sampled five times successively and processed the same way as the rest of the photosynthetic pigment samples in this study. The results had a SD of less than 5% from the mean for all photosynthetic pigments examined **(Table S1.1)**. HPLC analysis employed an internal control of including the same sample in two different quantities to ensure precision in pigment quantification, which scaled accordingly.

1.4 RESULTS

1.4.1 OD750/Cell Density Difference Between the Experiments

As detailed in 1.3.1, two distinct experiments were performed, in which cultures were equilibrated to a 12h:12h light-dark cycle under a sinusoidal light and temperature regime **(Fig. 1.1)** by nightly dilutions. The dilutions were based on cell counts, kept consistent throughout both experiments: for Exp. 1, the cultures were diluted to 3.1-4.2x10^5 cells·mL^{-1}; for Exp. 2, the post-dilution culture densities were 2.4-4.5x10^5 cells·mL^{-1}. This did not, however, correspond to consistency in OD750. Whereas post-dilution OD750 range for Exp. 1 was 0.5-1, the post-dilution OD750 values were twice that during Exp. 2 **(Fig. S1.1)**.

1.4.2 Cell Cycle Synchronization

In diatoms, cell cycle progression typically consists of the following phases: G1 (growth, when the cell size increases), S (chromosomal replication), and what has been defined as a combined G2/M phase because a true pause in progression (G2) may or may not occur prior to M phase (mitosis). Additionally, chloroplasts and mitochondria must replicate prior to mitosis. In diatoms, the generation of two daughter cells via mitosis occurs, but the cells do not separate until

later, after the silica cell wall is complete. Thus, cell number in a culture increases after a peak in the population of cells in G2/M phase. It should be noted that synchronization encompasses cell-to-cell variation in the rate of cell cycle progression, and unless the cells are in a prolonged stage of the cell cycle (e. g. G1), the percentage of cells in a given phase is never 100%.

Based on evaluation of cell cycle stages, cell cycle synchronization occurred in both experiments. Better synchronization was attained during Exp. 2, with >95% of the cells in G1 at dawn and also after the majority of the cells underwent division, compared to 66% at dawn and 75-80% after the majority divided in Exp. 1 **(Fig. 1.2)**. In Exp. 1, the peak percentage of cells in S phase occurred at 11:00 h, and in Exp. 2 at 11-12:00 h. The G2/M maximum was at 12-13:00 h for Exp. 1, and 13:00 h for Exp. 2. Increase in cell concentration in both experiments began after most cells in the culture entered G2/M. In both experiments, there was also a G2/M shoulder at the beginning of the dark period, and a subpopulation of cells that divided thereafter. The starting cell concentration was very similar between the experiments, but approximately 20% less at the end of Exp. 1 than Exp. 2 **(Fig. 1.2)**.

1.4.3 OD750 Dynamics

1.4.3.1 Lack of Correspondence Between OD750 and Cell Concentration

As described in 1.3.4, OD750 can be used as proxy for AFDW, a measure of biomass, for actively growing *C. cryptica* under highly diverse conditions. In both experiments, changes in OD750 and cell concentration followed generally similar patterns throughout the intensive sampling period, but OD750, and therefore biomass, started to increase several hours before cell concentration **(Fig. 1.3A, B)**. The rate of OD750 increase slowed significantly toward the end of the light period, but some increase during the dark period was observed in both experiments. The

starting OD750 for Exp. 2 was approximately twice that of Exp. 1; OD750 approximately doubled during the Exp. 2 intensive sampling period, and almost tripled for Exp. 1 **(Fig. 1.3A, B)**. If cell concentration and the corresponding OD750 values are plotted sequentially, there is a clear initial period of biomass increase only, followed by increase in cell concentration along with biomass **(Fig. 1.3C, D)**.

1.4.3.2 OD750 Increase Rate Correlates with Light Energy Availability

Optical density increase rate as a function of light energy availability was plotted as the change in OD750 per hour, normalized to culture density, versus average incident light intensity per hour during the light period. The data indicate a direct but non-linear correlation between optical density increase rate and incident light intensity **(Fig. 1.4)**. For both experiments, the rate of optical density increase per cell relative to incident light intensity trended higher during the second half of the day **(Fig. 1.4)**.

1.4.3.3 Inverse Relationship Between Daily OD750 Increase and Starting OD750

Plotting OD750 at the end relative to the beginning of the light period according to the specific growth equation [(ln($OD_{evening}$-$OD750_{morning}$))/time)] showed that the lower the starting OD750 (i.e., biomass) was at dawn, the more gain in OD750 occurred during the day, even at very low OD750 values **(Fig. 1.5)**. The observation that starting OD750 was higher for the intensive sampling day of Exp. 2 but gain in OD750 by the end of the intensive sampling day was greater during Exp. 1 is consistent with this trend **(Fig. 1.3)**.

1.4.3.4 OD750 Decrease at Dawn

A dip in OD750 within the first 2 hours of the light period was consistently observed for both experiments during the equilibration period and intensive sampling **(Fig. 1.6)**. Due to the sinusoidal nature of the light regime, irradiance increased substantially immediately after dawn, and had reached half of the maximal light intensity by the time OD750 began increasing after the dip (1000 µmol photons m^{-2} s^{-1} at 8:00 h) **(Figs. 1.1, 1.6).**

1.4.3.5 OD750 Decrease at Night

The net average decrease in OD750 at night, signifying biomass loss, was 5.7±1.1% for Exp. 1 and 5.6±1.6% for Exp. 2 during the days the cultures were diluted **(Table S1.2)**. The OD750 decrease at night was higher prior to the initiation of dilutions, decreasing with each post-inoculation day in Exp. 2. No trend could be observed for Exp. 1, as only one day after inoculation was required to reach the desired cell concentration after which dilutions were initiated **(Figs. S1.1, S1.4, Table S1.2)**.

1.4.4 Cellular TAG Dynamics

In both experiments, there was a slight decrease in cellular TAG levels after dawn, followed by a steep increase until the S-phase peak **(Fig. 1.7)**. In Exp. 1, the increase started at 7:00 h, one hour into the light period, and peaked at 11:00 h. In Exp. 2, TAG levels increased sharply from 11:00 h to 12:00 h. A mid-day dip in cellular TAG levels followed, corresponding to the G2/M peak, a period when membrane synthesis associated with cell division is expected. As most cells completed division, the cellular TAG reserves were replenished, before declining several hours into the dark

period, after 21:00 h. Prior to the decline, a significant increase in cellular TAG after dusk was observed in Exp. 2. The decline continued until 5:00 h, the end of the intensive sampling period for Exp. 2; for Exp. 1, cellular TAG increased again after 1:00 h **(Fig. 1.7)**.

1.4.5 Photosynthetic Pigment Dynamics

1.4.5.1 Changes in Cellular Photosynthetic Pigment Content During Intensive Sampling Period

The major *C. cryptica* photosynthetic pigments are chlorophyll a (Chl a), fucoxanthin (Fx), chlorophyll c (Chl c), beta-carotene (β-car), diadinoxanthin (Ddx), and diatoxanthin (Dtx), with the first three being light-harvesting and the last three photoprotective. Chl a is the major light-harvesting pigment, found in photosystem reaction centers as well as in photoantenna complexes with the accessory photopigments Fx and Chl c, of which Fx is more abundant. Ddx is converted to Dtx by de-epoxidation in response to light-induced stress, and the reaction is reversed during recovery. Because Ddx and Dtx are able to interconvert rapidly, we examined their abundance as a pool as well as separately. β-car also plays an important role in photoprotection, but does not participate in the aforementioned cycle [Croce and van Amerongen 2014, Kuczynska et al. 2105].

Photosynthetic pigment content per cell was measured throughout the intensive sampling period for both experiments **(Figs. 1.8, S1.5)**. All of the pigments had a daytime peak, and all except for Ddx and Dtx fluctuated during dawn, dusk, and early dark period, with fluctuations in Chl a content being the highest in amplitude **(Figs. 1.8, 1.9, S1.5)**. For both experiments, Chl a steadily increased after fluctuating at dawn, maximizing at noon **(Fig. 1.8A, B)**. During Exp. 1, Chl a levels plateaued thereafter, prior to onset of dusk fluctuations. For Exp. 2, the Chl a peak at noon was followed by a decrease prior to fluctuations at dusk. The overall trend for Fx in Exp. 1 was an increase from dawn to dusk **(Fig. 1.8C, D)**. During Exp. 2, Fx peaked at noon, with a similar

distribution on either side of the peak between dawn and dusk. The patterns of change in Chl a and Fx were similar to each other within each experiment, but with some differences (1.5.5) **(Fig. 1.9)**. The Ddx+Dtx pool abundance, determined mainly by the more abundant Ddx, peaked at noon during both experiments, with a similar distribution on either side of the peak during the light period **(Fig. 1.8E, F)**. Ddx followed that pattern of accumulation closely, whereas Dtx followed a different pattern (1.4.5.3) **(Fig. 1.10)**. Patterns of change in Chl c and β-car content paralleled that of Fx closely for both experiments **(Fig. 1.9)**. After initial fluctuation in the beginning of the dark period, there was a steady decline in Chl a, Fx, Chl c, and β-car abundance in the dark for the remainder of both experiments, whereas the Ddx+Dtx content remained relatively stable after an initial dip followed by a slight increase at the beginning of the dark period. For Exp. 1, the cellular content of all measured pigments was similar between the beginning and the end of the intensive sampling period, with Fx and Chl c slightly higher towards the end. For Exp. 2, all measured pigments were lower at the end than at the beginning of intensive sampling **(Figs. 1.8-1.10, S1.5)**. The amount of each individual pigment was higher during Exp. 1 than Exp. 2, with the exception of Dtx, which was present, on average, in the same quantity during both experiments **(Table 1.1)**.

The observed fluctuations in the levels of different pigments varied in both amplitude and timing, as demonstrated by co-plotting normalized pigment levels **(Figs. 1.9, S1.7)**. Although only one sample per time point was taken due to sampling time limitations, these differences support the validity of the observed fluctuations, arguing against them being caused by sampling or processing error. Furthermore, sample to sample variation in photosynthetic pigments was ascertained to be too low to account for all of the observed fluctuations (1.3.8), with a SD of 0.9-4.9% of the mean of five consecutively obtained control samples for all assayed pigments **(Table S1.1)**. By contrast, the SD of the mean of the first five time points taken during Exp. 1 was 6.8-8.9%

for all examined pigments, and 3.6-15.4% for the first five samples obtained during Exp. 2, with each individual pigment varying more during the experiments than the control.

1.4.5.2 Photosynthetic Pigment Abundance, Cell Cycle Progression, and Changes in Light Intensity

Examination of data in **Figs. 1.8, 1.9** and **S1.5** suggests a correlation between changes in photosynthetic pigment content per cell and changes in irradiance and/or cell cycle progression. The difference in the extent of synchrony between the two experiments allows for discernment of the potential influence of the two variables. 66% of the Exp. 1 culture and >95% of the Exp. 2 culture were similarly synchronized at dawn on the intensive sampling days (1.4.2). The daytime peaks in Chl a, Fx, Chl c, and β-car content occurred around the S phase to G2/M transition and were more protracted during Exp. 1 than in Exp. 2, along with a less complete cell cycle synchronization. This suggests that although changes in the cellular abundance of these pigments during Exp. 2 could be interpreted as following the changes in irradiance, the lower extent of synchrony in Exp. 1 reveals cell cycle as a contributing factor. By contrast, Ddx+Dtx in both experiments peaked at 12:00 h along with the light intensity and had an equal distribution on both sides of the peak during the light period, following the sinusoidal light intensity regime **(Fig. 1.8E-H)** and without observable differences due to varied extents of synchrony and therefore cell cycle progression. Additionally, the Ddx+Dtx levels remained relatively stable in the dark, whereas the other measured pigments gradually decreased. It appears that the abundance of Chl a, Fx, Chl c, and β-car may be influenced by the timing of cell cycle progression, whereas the Ddx+Dtx content may follow the changes in light intensity only **(Figs. 1.8, 1.9, S1.5)**.

1.4.5.3 Dtx Dynamics and Ddx+Dtx De-epoxidation State

Dtx abundance, along with the de-epoxidation state of the Ddx+Dtx pool (Dtx/(Ddx+Dtx)), which increase in response to light-induced stress, decreased immediately upon dawn. These parameters did not maximize along with light irradiance, unlike the Ddx+Dtx pool abundance **(Figs. 1.8, 1.10)**. Rather, the daytime peak occurred before noon and was more protracted in Exp. 1 than in Exp. 2, which had a tighter peak after noon. Additionally, the de-epoxidation state increased sharply immediately following the onset of the dark period during both experiments, then returned to approximately the pre-dawn values by the end of the intensive sampling period **(Fig. 1.10)**.

1.4.5.4 Ratio Dynamics of Chl a to Accessory Photopigments (Fx and Chl c)

The ratios of Chl a to the accessory photopigments Fx and Chl c paralleled each other very closely and were slightly higher for Exp. 1 **(Table 1.1)**. The overall trend for the ratios was an increase during the day, with 1.6-1.9X differences between pre-dawn and maximum values **(Fig. 1.11)**. After the evening fluctuations, the ratios for both experiments gradually decreased to approximately the pre-dawn levels observed at the beginning of sampling **(Fig. 1.11)**.

1.5 DISCUSSION

1.5.1 Overview

In this study, we strove to examine general diel trends in physiological and metabolic rhythms of the diatom *C. cryptica* in controlled simulated outdoor conditions using sinusoidal variation in light intensity and temperature. Although other environmental variables are a factor in outdoor cultivation, their transient and/or irregular occurrence can obfuscate overall larger-scale

trends, and elucidating such trends was the goal of this study. We performed two separate experiments, attempting to adapt the cultures to the same cultivation conditions prior to intensive sampling by diluting them each night for a week once they reached the desired cell density, with some differences in the timing of dilutions between the two experiments. There was variance between the cultures, potentially arising from the differences in the state and density of the inoculum and the details of pre-adaptation (1.3.1, 1.4.1), but the general observed trends were consistent. The differences found despite the cultivation conditions being carefully controlled and reproduced between the two experiments are a testament to the variability inherent to working with biological systems, and are worth documenting and appreciating as such.

1.5.2 Cell Cycle Synchronization

Synchronous division has been typically assessed by tracking changes in culture density and/or division rates [Anderson and Sweeney 1977, Bonnefond et al. 2016, Chauton et al. 2013, Chisholm et al. 1980, Darley and Volcani 1971, Eppley et al. 1967, Hoogenhout 1963, Jallet et al. 2016, Kohata and Watanabe 1989, Nelson and Brand 1979, Ogbonna and Tanaka 1996, Owens et al. 1980, Post et al. 1984, Ragni and d'Alcala 2007, Sukenik and Carmeli 1990, Varum et al. 1986, Williamson 1980]. To our knowledge, this is the first study to track cell cycle progression and relate it to diel physiological and metabolic changes in light-dark synchronized microalgal cultures, allowing for a more detailed understanding of the observed phenomena. In our experiments, *C. cryptica* synchronized to predominantly divide during the light period **(Fig. 1.2)**. A greater extent of cell cycle synchronization was attained in Exp. 2 than in Exp. 1; this may have been due to a difference in factors that we did not specifically investigate, such as length of the pre-adaptation period and the timing and consistency of dilutions (1.3.1). In addition to the main cell division event

in the middle of the light period **(Fig. 1.2)**, there is indication that a subset of cells (approximately 20%) was dividing earlier during the light period, then amassing enough energy to divide again in the dark. This is evidenced by a G2/M shoulder at the beginning of the dark period, the timing of which is the same for both experiments despite the synchronization differences **(Fig. 1.2)**. Since culture density more than doubled during the intensive sampling days for both experiments **(Fig. 1.2)**, the explanation that a subpopulation of cells had divided more than once in a 24-hour period appears more likely than the notion that the late-dividing subpopulation only divided once and substantially later than the majority of the cells in the culture. Incomplete synchronies and the possibility of multiple synchronous subpopulations arising from growth regime selection have been observed in other studies and are discussed in Hoogenhout [1963].

1.5.3 Biomass Dynamics

OD750 is a convenient and popular method for monitoring microalgal growth, as the 750 nm wavelength is not absorbed by microalgal pigments, allowing for light scattering alone to be measured [Chioccioli et al. 2014]. Due to the susceptibility of light scattering to influence by multiple variables such as culture density, cell size, shape, and composition, it has been documented that there is not always a linear relationship between OD750 and biomass measured as AFDW [Chioccioli et al. 2014, Edmundson and Huesemann 2015]. However, we found a robust correlation between OD750 and AFDW for actively growing *C. cryptica* (1.3.4) **(Fig. S1.3)**. Our ePBR experiments were carried out in the same growth medium (ASW) that was used in the raceway pond experiments from which that conclusion had been derived, using similar sinusoidal light and temperature parameters. Although the OD750 curves do not show perfectly smooth responses, they do not follow the more drastic dynamics of cell cycle changes (including cell division),

consistent with OD750 in our study predominantly being related to biomass and not the other aforementioned variables **(Fig. 1.3A, B)**. This has enabled us to use OD750 to examine biomass dynamics in our cultures with a previously unreported (to our knowledge) 10-minute interval resolution.

In our study, OD750 increased for the majority of the light period, except for dawn **(Figs. 1.3, 1.6)**. Culture density, however, increased only in the second half of the light period **(Fig. 1.3)**, suggesting that cells initially accumulated biomass in preparation for division and continued to accumulate it throughout the light period. The rate of biomass accumulation increased with light intensity **(Fig. 1.4)**, consistent with energy input correlating with biomass output, and also suggesting that there was no light saturation or substantial photoinhibition [Ho et al. 2012, Sukenik et al. 1987], despite the irradiance being high for the majority of the light period **(Fig. 1.1)**. This may be attributed to the combination of high culture density and the deep and narrow shape of the ePBR [Lucker et al. 2014] resulting in the average irradiance experienced by the culture being substantially lower than the incident light, and below saturating light intensity. The rate of biomass accumulation per cell relative to incident light intensity trended higher during the second half of the day **(Fig. 1.4)**. Since culture density during the second half of the day was higher than it was during the first half, the culture was more productive overall. By then, most cells would have already divided, and thus would not be expending carbon on division, which could allow for more of it to be stored. Other factors that differed between the two halves of the day were temperature **(Fig. 1.1)** and the amount of pigments and thus shading in the culture **(Fig. 1.12)**, both higher during the second half. The best fit curves are not linear **(Fig. 1.4)**, and suggest that the efficiency of light to biomass conversion improves with increasing light intensity, for reasons that are unclear.

We observed an inverse relationship between the starting OD750 at dawn and the subsequent OD750 increase during the day, with lower starting OD750 values corresponding to

greater OD750 increase **(Fig. 1.5)**. A similar relationship had been noted previously [Grima et al. 1995, Michels et al. 2014]. Consistent with this trend, the starting OD750 for the Exp. 1 intensive sampling day was lower than that for Exp. 2, while the gain in biomass by the end of the day was greater **(Fig. 1.3)**. This observation may be explained by higher OD750 values corresponding to lower light penetrance into the culture and therefore light limitation, even at low densities **(Fig. 1.5)**. Less access to light would lead to less carbon fixation relative to respiration, and therefore less ability to store compounds for energy. This ties in with the observation that the rate of biomass accumulation increases with light intensity **(Fig. 1.4)**.

As mentioned above, biomass did not increase during the beginning of the light period, which was also reported by Grima et al. [1995] for the diatom *Phaeodactylum tricornutum* cultivated outdoors. We observed a dip in OD750 during the first 1-1.5 hours of the light period, during which light intensity increased substantially **(Figs. 1.1, 1.6)**. Biomass did not start to increase for the day until the light intensity reached nearly half of the maximum (1000 μmol photons m^{-2} s^{-1} at 8:00 h), suggesting that lack of light availability during that time period was not accountable for the observed phenomenon. There was also a dip in TAG around then, though the timing didn't strictly correlate **(Fig. 1.7)**. It is possible that energy-requiring processes began when light was sensed, but not enough energy was initially obtainable from the environment to power them and supersede respiration. Thus, stored energy, including TAG, would have needed to be mobilized, resulting in a reduction in TAG and biomass.

Biomass gains made by microalgal cultures during the day are offset by losses at night due to nighttime energy-requiring processes and the accompanying respiration [Burris 1977, Cuhel and Lean 1987, Edmundson and Huesemann 2015, Geider and Osborne 1989]. The extent of the losses depends on the species, physiological state of the culture, and environmental conditions. Generally, the higher the metabolic rate, which correlates with greater light exposure during the day, lower

culture density, exponential growth, favorable cultivation conditions, and higher temperature at night, the greater the nighttime losses will be [Bonnefond et al. 2016, Edmundson and Huesemann 2015, Geider and Osborne 1989, Grima et al. 1995, Grobbelaar and Soeder 1985, Hu et al. 1998, Le Borgne and Pruvost 2013, Michels et al. 2014, Ogbonna and Tanaka 1996]. In some cases, losses of more than 30% have been documented [Guterman et al. 1988, Hu et al. 1998, Torzillo et al. 1991]. In our study, OD750 as a proxy for biomass gained during the day was lower and the nighttime losses higher immediately post-inoculation than after the cultures had some time to adapt, with the effect being more salient for Exp. 2 **(Table S1.2, Figs. S1.1, S1.4)**. This might be explained by the metabolic states of the inocula, higher initial biomass in the case of Exp. 2 resulting in less potential for gain during the day, and initial lack of synchrony resulting in more energy expended at night due to division processes. During the intensive sampling periods, nighttime losses were 5.7±1.1% and 5.6±1.6% of OD750 gained during the day for Exp. 1 and Exp. 2, respectively **(Table S1.2, Figs. S1.1, S1.4)**. Since the extent of biomass loss at night has a substantial impact on the overall productivity, it is imperative in a production scenario to take it into account and minimize by strain selection and optimization of cultivation conditions. In addition to lowering the temperature at night, minimizing culture mixing at night can also lessen respiratory losses by decreasing dissolved oxygen, though this approach may negatively impact nighttime biosynthetic processes which may be important to overall productivity [Edmundson and Huesemann 2015, Ogbonna and Tanaka 1996]. Additionally, it is important to optimize biomass concentration so as to maximize the ratio of light utilization during the day to respiratory losses at night [Michels et al. 2014].

1.5.4 TAG Dynamics

TAG, of commercial interest for biofuel production, has a variety of functions in actively growing microalgal cultures and tends to hyperaccumulate in mostly cytoplasmic lipid droplets during periods of environmental stress such as nutrient limitation [Goodman 2008, Hu et al. 2008, Maeda et al. 2017, Roessler 1990, Solovchenko 2011, Thompson 1996, Zehmer et al. 2009]. In diatoms as well as other microalgae, TAG is made in smaller amounts during favorable growth conditions, typically accumulating during the day when photosynthesis occurs, and depleting at night to be used for energy and biosynthesis of other molecules [Anderson and Sweeney 1977, Bonnefond et al. 2016, Chauton et al 2013, Jallet et al. 2016, Lacour et al. 2012, Sukenik and Carmeli 1990].

Consistent with the aforementioned observations, cellular TAG levels increased toward the middle of the light period and remained high until 3 hours into the dark period during both of our experiments, suggesting that some of the carbon fixed during the day was stored as TAG, then broken down for energy and/or repartitioned into other types of molecules at night. The mid-day dip in cellular TAG levels corresponded to the time when the majority of the cells in the culture were in G2/M during both experiments **(Fig. 1.7)**, likely because TAG was used for membrane lipid biosynthesis [Athenstaedt and Daum 2006, Kurat et al. 2009, Solovchenko 2011]. An alternate/additional explanation may have been that the dip occurred at the time that cell concentration began to increase **(Fig. 1.2)**, and as cells divided, their TAG content was split between the daughter cells, giving a temporary appearance of decreased cellular TAG, while the total TAG, defined as the combined TAG content of all the cells in the culture, remained unchanged or continued to increase. However, the total TAG decreased at that time as well **(Fig. S1.6)**, supporting the hypothesis that some of the stored TAG was used for cell division.

TAG levels increased after dark for both experiments, and more substantially during Exp. 2 **(Fig. 1.7).** Since no new carbon was being fixed at that time, the nighttime increase in TAG may be attributed to repartitioning of carbon from other molecules. The nighttime increase was followed by a decrease to approximately dawn levels, potentially due to carbon repartitioning and/or respiratory losses **(Fig. 1.7)**. Consistency in TAG levels at dawn on consecutive days despite the transient daytime increase has also been observed in other actively growing microalgal cultures entrained to a light-dark regime, suggesting that such cultures transiently accumulate TAG to be actively used on a daily basis [Chauton 2013 and refs within, Jallet et al. 2016, Lacour et al. 2012, Sukenik and Carmeli 1990].

1.5.5 Photosynthetic Pigment Dynamics

Most previous studies documenting diel changes in microalgal photosynthetic pigment abundance have been carried out in constant temperature using a square-wave light regime [Kohata and Watanabe 1988, Kohata and Watanabe 1989, Owens et al. 1980, Post et al. 1984, Ragni and d'Alcala 2007]. Recently, two groups utilized a sinusoidal light regime, Jallet et al. [2016] examining the Chl a dynamics of the diatom *P. tricornutum* in constant temperature, and Bonnefond et al. [2016] using a sinusoidal temperature regime to examine photosynthetic pigment dynamics of the chlorophyte *Dunaliella salina*. As far as we are aware, this is the first report of photosynthetic pigment dynamics in a diatom using both sinusoidal light and temperature, with examination of diel changes in pigmentation at higher temporal resolution than previously published for any microalgal species. Specifically, whereas other groups typically sampled for pigments every few hours, we sampled as frequently as every 15 minutes during the dawn and dusk

periods when light intensity was changing rapidly **(Fig. 1.1)**. This is also the first report, to our knowledge, to overlay the photosynthetic pigment dynamics data with cell cycle progression.

As seen in other studies [Jallet et al. 2016, Post et al. 1984, Ragni and d'Alcala 2007], cellular photosynthetic pigment content increased during the day and decreased by the following dawn during both of our experiments, returning to approximately pre-dawn levels of the previous day **(Figs. 1.8-1.10, S1.5)**. As discussed below, Dtx followed a unique pattern relative to the other measured pigments **(Fig. 1.10)**, but the abundance of the Ddx+Dtx pool exhibited the general trend of increase during the day followed by a decrease, because Ddx was the more abundant form of the two interconvertible pigments **(Figs. 1.8E-H, 1.10A, B).** There appears to be a difference between the dynamics of Ddx+Dtx and the other photosynthetic pigments (light-harvesting pigments and β-car), potentially indicating differences in regulation (1.4.5.2). Ddx+Dtx abundance appears to be regulated by light intensity and not affected by cell cycle progression, whereas the other photosynthetic pigments appear to be influenced by cell cycle progression and not strictly by light intensity. Ragni and d'Alcala (2007) obtained data for diel photosynthetic pigment variation in the diatom *P. tricornutum* that agree with our observations, although sampling for pigments at a lower temporal resolution and using a square-wave light regime. In their study, the light-harvesting pigments and β-car gradually increased throughout the light period and peaked around dusk, immediately before the majority of the culture underwent cell division, consistent with the notion that the timing of accumulation of those pigments relates to cell cycle progression. They did not observe Dtx production under their experimental conditions, however Ddx maximized much earlier than the other pigments, closer to the beginning of the light period, and remained relatively stable before decreasing in the dark. The nearly square response to the square-wave light regime, in contrast to the sinusoidal-like response to the sinusoidal light regime in our study **(Fig. 1.8G, H)**, further supports the hypothesis that the Ddx+Dtx pool is regulated directly by light intensity. Owens

et al. [1980] found that the abundance of Chl a and Chl c exhibited rhythms that followed the light-dark cycle in the diatom *Skeletonema costatum* during asynchronous growth and stationary phase, with accumulation in the light and degradation in the dark. They suggested that Chl a and Chl c accumulation may be influenced by light-dark cycling as well as cell cycle progression, with synchronized division affecting the timing of photosynthetic pigment accumulation in cultures. The behavior of other pigments was not assessed. Considering the data obtained in our study and by Ragni and d'Alcala [2007], we agree with the suggestion, and extend it to potentially include Fx and β-car as well. Increase in photosynthetic pigment content per cell in preparation for division could be due to the need to populate newly divided chloroplasts. If the division takes place in daytime, increased shading may also play a role. The extent to which the shape of the light intensity curve influences timing of the accumulation of the above pigments is yet to be determined.

Sampling for photosynthetic pigment abundance with high resolution has allowed us to observe the extent to which different pigments co-varied with each other. As discussed above, the overall trends were similar for the light-harvesting pigments and β-car. However, Fx, β-car, and Chl c co-varied more closely with each other than with Chl a **(Fig. 1.9)**. The observation that Chl c dynamics were more similar to Fx than to its closer chemical relative Chl a may relate to the intricacies of the molecular architecture of light-harvesting complexes and their dynamics. β-car is an earlier intermediate of the carotenoid biosynthesis pathway, which has Fx and Ddx+Dtx as the end products, thus the observed close co-variation of β-car with Fx is not surprising. Much remains to be understood about the pathway in diatoms, including the sequence of the final steps leading to the biosynthesis of Ddx and Fx and the enzymes responsible for their catalysis. One of the leading hypotheses is that Ddx serves as a precursor for both Dtx and Fx [Kuczynska et al. 2015], which makes the apparently separate regulation of the Ddx+Dtx pool from the rest of the carotenoid biosynthesis pathway, including the potentially downstream Fx, very interesting. The distinction

between Ddx+Dtx and Fx dynamics despite the possible precursor/product relationship suggests additional regulatory mechanics. This finding must be taken into consideration during further investigation of the pathway and its regulation.

Substantial fluctuations in the levels of Chl a, Fx, Chl c, and β-car, but not Ddx+Dtx, were observed during the dawn and dusk periods of rapid light intensity changes, providing further evidence for separate regulatory mechanisms **(Figs. 1.8-1.10, S1.5)**. Photosynthetic pigment-level adjustments to abrupt and unanticipated changes in growth irradiance are known to happen on the scale of hours to days [e.g., Nymark et al. 2009, Post et al. 1984], but there are, to our knowledge, no high-resolution data available for photosynthetic pigment dynamics during anticipated gradual changes in light intensity. As described by Post et al. (1984), on longer time scales, diel rhythms in cellular Chl a abundance relative to changes in cellular Chl a content due to light intensity adaptation are kinetically distinct and can be mathematically resolved. They observed an increase or decrease in average daily Chl a content when diatom cultures grown using a square-wave light regime were shifted to lower or higher growth irradiance, respectively. At the same time, daily rhythms in Chl a abundance persisted. They posited that the fine control over cellular Chl a content may be explained by adjustments between biosynthetic and degradation rates. We suggest that the same may be the case for Chl c, Fx, and β-car, pending direct experimental evidence. Given what is known about the kinetics of pigmentation adjustment to irradiance changes, it is possible but unlikely that the dawn and dusk fluctuations we observed were due to dynamic adjustments of the photosynthetic apparatus to the rapidly changing light intensity during dawn and dusk, overcompensating in each direction while aiming for a moving target Rather, we postulate that general remodeling of light-harvesting pigment-protein complexes from a daytime to a nighttime state might take place at those times. The continuation of the fluctuations into the beginning of the dark period is supportive of that hypothesis. Monitoring the abundance of relevant proteins at

those times, as well as assessing the dawn, dusk, and early dark period photosynthetic pigment and protein response at high resolution utilizing a square-wave light regime, when the change from no illumination to maximum light intensity is abrupt rather than gradual, but administered at regular intervals, would help distinguish between the contributions of entrained diel periodicity and spontaneous light intensity adaptation to the fluctuations we observed at dawn and dusk under a sinusoidal light regime.

In diatoms, there is a direct correlation between Dtx concentration and non-photochemical quenching (NPQ), a photoprotective mechanism that allows for dissipation of excess light energy as heat [Kuczynska et al. 2015, Lepetit 2012]. Although Dtx was much less abundant than Ddx during the course of both experiments **(Fig. 1.10A, B)**, it exhibited reproducible patterns of accumulation, which can also be conceptualized in relation to the Ddx+Dtx pool as its de-epoxidation state, Dtx/(Ddx+Dtx) **(Fig. 1.10C, D)**. There was a de-epoxidation state peak during the light period that did not correspond to the timing of maximum Ddx+ Dtx abundance, which occurred at noon for both experiments **(Fig. 1.8E-H)**, perhaps due to correlation with cell cycle progression in addition to light intensity. A general trend of an increase in Dtx abundance towards the middle of the light period, followed by a decrease towards dusk was present, but not in the nearly bell-shaped curve of the Ddx+Dtx pool abundance, which the more copious Ddx paralleled more closely **(Figs. 1.8E-H, 10)**. Relatively low de-epoxidation state maxima and lack of very close correlation with light intensity can be attributed to the fact that the cultures were fairly dense, so a substantial proportion of cells would not be receiving a stress-inducing amount of irradiation at any given time. What may have contributed to the timing of Ddx de-epoxidation and the presumable development of NPQ during the light period is the separation of replicated chloroplasts and/or daughter cells as the result of cell cycle progression, resulting in them losing localized shading due to the package effect (reduced *in vivo* light absorption efficiency due to the geometric organizational properties of

photosynthetic pigments in intact cells) [Geider and Osborne 1987]. The observation that the de-epoxidation state had an earlier and more protracted daytime peak in Exp. 1 is consistent with this hypothesis, as cell division started earlier than in Exp. 2 due to the lesser extent of synchronization **(Fig. 1.10C, D)**. The unanticipated influence of other cell division-related processes on the de-epoxidation state cannot be ruled out, but we are not aware of studies on the subject. Epoxidation of Dtx in low light after a period of high light exposure is known to occur on the scale of minutes, as it is necessary in diatoms for resuming light harvesting activity [Goss et al. 2006, Grouneva et al. 2009], and the decrease of the de-epoxidation state by dusk is consistent with that. Another de-epoxidation state peak occurred immediately after the onset of darkness in both experiments. This can be explained by a chlororespiratory transthylakoid pH gradient that activates the Ddx de-epoxidase, leading to the accumulation of Dtx and NPQ in the dark [Jakob et al. 1999, Jakob et al. 2001]. Although the de-epoxidation state appears higher during the nighttime peak than during the day, the cellular abundance of Dtx is actually lower at that time, as is the Ddx+Dtx pool abundance; it is the proportion of the pool that is de-epoxidized that creates this appearance, not an increased absolute abundance of cellular Dtx **(Fig. 1.10)**. After the initial nighttime peak, the de-epoxidation state decreased to pre-dawn levels. To our knowledge, no previous study was designed to document this phenomenon, and the exact causes for it remain to be elucidated. The observed rapid decrease in the de-epoxidation state upon dawn can be attributed to the inhibition of chlororespiration by the presence of light [Peltier et al. 1987] and the need for the system to transition from a dissipative to a light-harvesting state to take advantage of the available irradiance. Jakob et al. [1999] also observed epoxidation of night-accumulated Dtx under low irradiance at dawn in *P. tricornutum* grown using a light-dark cycle with an exponential illumination regime. It should be noted that Dtx epoxidation proceeds significantly faster in low light than in the dark, due

to the scarcity of NADPH, which is a co-factor for the Dtx epoxidase, in darkness [Goss et al. 2006, Grouneva et al. 2009].

Changes in the ratios of Chl a to the accessory photopigments Fx and Chl c may indicate changes in the composition/structure of the photosynthetic apparatus. The observed ratio fluctuations at dawn, dusk, and the beginning of the dark period **(Fig. 1.11)** indicate a dynamic adjustment of the pigments during which Chl a varied distinctly from Fx and Chl c, which varied similarly to each other **(Fig. 1.9)**. Other groups have demonstrated that these ratios in diatoms remain relatively constant regardless of growth irradiance [Lepetit et al. 2012]; in accordance with those observations, and because the daytime increase in the ratios during our study did not exhibit a trend that would be consistent with them changing with light intensity, cell cycle progression, or increase in culture density, we suggest that the day-night differences may be due to diel changes in light-harvesting complex assembly, as discussed above. Ragni and d'Alcala (2007) did not observe periodic variations in the ratios during their study; we note the discrepancy, but do not attempt to explain it, due to the scarcity of this type of data obtained in controlled laboratory conditions to date. There are data from diatom-dominated field samples, however, that also indicate larger changes in Chl a content during the day in comparison to Chl c and carotenoids [Yentsch and Scagel 1958].

1.5.6 Product Accumulation and Harvest Timing

In a production scenario where microalgae experience a light-dark cycle, such as outdoors, it may be possible to take advantage of the predictable timing of various processes by deducing the best times to harvest for biomass and/or other products. For example, in our set-up with *C. cryptica*, biomass increased throughout the day, slowing around 17:00 h **(Fig. 1.3A, B)**. Harvesting at

that time, after the majority of biomass gain had been completed for the day, rather than at 8:00 h (i.e., at the beginning of a typical work-day), would have approximately doubled the yield in Exp. 2, and tripled it in Exp. 1 **(Table S1.3)**. Extending the harvesting time into the dark period would not have diminished the potential yield in our study, because the majority of cell division occurred during the day and nighttime respiratory losses were minimal. However, if we were working with a species that divided at night, the timing would work out differently. Biomass would likely still maximize towards the end of the light period, unless there was an external source of carbon that could be consumed heterotrophically without light energy. Cell division, however, would ensue in the dark, and the energy required for associated processes would result in a reduction of biomass available for harvest. In that scenario, it would be advantageous to harvest for biomass at the end of the light period and prior to the onset of cell division. Thus, when assessing the biomass productivity of a given strain under given conditions, it should be noted that a nocturnally dividing strain would use more energy at night than a strain synchronized to divide during the day [Grobbelaar and Soeder 1985]. Therefore, netbiomass change in a 24-hour period, rather than daytime gain or nighttime loss alone, needs to be assessed when determining a strain's overall biomass productivity in given culture conditions.

The optimal time to harvest for TAG in our study, determined by the maximal cell concentration with the highest cellular TAG levels, was 21:00 h for both experiments **(Fig. 1.12A, B)**. Harvesting at that time would have improved yield approximately 4-6-fold compared to 8:00 h for both experiments **(Table S1.3)**. In a culture dividing at night, TAG levels would still be maximal at the end of the light period, and it would thus be beneficial to harvest prior to onset of energy-requiring division processes.

Microalgal pigments are also of commercial interest for a variety of applications such as functional nutrition and cosmetics [Khanra et al. 2018, Matos et al. 2017]. In *Cyclotella cryptica*, Chl

a, Fx, β-car, and Ddx+Dtx (quantified as a pool as they may interconvert during harvesting) are the pigments that accumulate to an appreciable amount and could be harvested, perhaps along with biomass and/or TAG, to serve as high value co-products that would help offset the costs of biofuel production. In our system, the optimal time to harvest for pigments spanned from dusk through the dark period for Fx and β-car, a narrower time frame during the dusk to night transition for Chl a, and an even more generous time frame for Ddx+Dtx, including more of the light period, with a dip during the Chl a optimal harvesting time **(Fig. 1.12C-H)**. The timing would be generally favorable for co-harvesting with either biomass or TAG, and, depending on the pigments, would improve yield approximately 1.5-3-fold (calculated as the ratio between the average pigment content for the preferred timeframe and 8:00 h) **(Table S1.3)**.

The abundance of other molecules of interest, such as proteins, carbohydrates, and polyunsaturated fatty acids (PUFAs), will also exhibit diel variation in light-dark synchronized microalgal cultures. The exact details of timing for maximum yield will vary with the specifics, such as the species used, the growth regime, production system metrics, and product(s) of interest. At present, many large-scale production systems employ time-consuming harvesting methods that might not allow for taking advantage of this type of information [Barros et al. 2015, Gerardo et al. 2015]. However, the gains in productivity achieved by an informed timing of harvest may potentially help offset the costs of faster harvesting methods or inspire the development of the latter.

1.6 SUMMARY AND CONCLUSIONS

Diel TAG, OD750 as a proxy for biomass, and photosynthetic pigment dynamics in the diatom *Cyclotella cryptica* grown using a sinusoidal light and temperature regime simulating what

microalgae experience outdoors were examined. As is typical in microalgal cultures grown on a light-dark regime, cell cycle synchronization occurred. The main cell division event was in the middle of the light period, preceded by an increase in biomass by a few hours. This observation emphasizes the fact that the two variables, which are often used as an assessment of culture growth, do not necessarily correlate. They would be even less parallel in a nocturnally dividing culture, which would gain biomass during sunlight hours and then lose a greater proportion of it at night due to respiratory requirements associated with cell division. Cellular TAG accumulated during the day and decreased at night. The observed daytime dip corresponded to the time when the majority of the population would require lipids for division-related membrane biosynthesis, suggesting that TAG may be serving as a precursor for membrane lipids. Differences in the extent of synchrony between the two experiments and patterns of accumulation of Chl a, Fx, Chl c, and β-car on a per cell basis suggest that cell cycle progression influences their cellular abundance. Care must be taken when normalizing other parameters to Chl a, which is common practice, as its abundance per cell is not constant. By contrast, the Ddx+Dtx pool abundance appeared to be dictated by irradiance and independent of cell cycle progression. This suggests a separate level of regulation for these end products of the carotenoid biosynthesis pathway from Fx, the other end product, and β-car, a pathway intermediate, and must be taken into account in further studies of the pathway in diatoms, which still remains poorly understood. Cell cycle progression did appear to affect the proportion of the Ddx+Dtx pool present as Dtx, which correlates with NPQ and is an indicator of irradiance-induced stress. As detailed in 1.5.5, this suggests that loss of local shading due to chloroplast division and/or separation may result in stress that triggers the photoprotective mechanism. Finally, our results indicate that timing is of importance when sampling light-dark synchronized cultures for study or harvesting them for production, as it will significantly affect the abundance of various cellular components and product yields. Because of accurately mimicking

outdoor light and temperature conditions, these lab-based experiments have conceptual relevance to outdoor cultivation. The species examined, *Cyclotella cryptica*, also has relevance to outdoor production, as it has been identified as a top production candidate [Traller et al. 2016].

1.7 ACKNOWLEDGEMENTS

Chapter 1, in full, has been submitted for publication. Gaidarenko, Olga; Sathoff, Corinne; Staub, Kenneth; Huesemann, Michael M.; Vernet, Maria; Hildebrand, Mark. "Timing is everything: diel metabolic and physiological changes in the diatom *Cyclotella cryptica* grown in simulated outdoor conditions." Ms. Gaidarenko was the principal author on this paper. We thank Dr. Susan Golden and her lab members for lending us the ePBR used for this work. We also thank Dr. Daniel Wangpraseurt for assistance with data interpolation. This work was supported by U.S. Dept. of Energy grant DE-SC0012556.

Table 1.1

Average cellular pigment content and ratios.

	Exp. 1 (pg/cell)	Exp. 2 (pg/cell)	Exp. 1:Exp. 2 Ratio
Chl a	4.34±1.15	3.58±0.66	1.21:1
Fx	1.14±0.10	1.04±0.08	1.09:1
Chl a/Fx	3.76±0.79	3.45±0.57	1.09:1
Chl c	0.17±0.02	0.15±0.01	1.13:1
Chl a/Chl c	24.95±5.35	23.38±4.13	1.07:1
β-car	0.10±0.02	0.08±0.01	1.25:1
Ddx	0.37±0.10	0.34±0.09	1.09:1
Dtx	0.05 ± 0.02	0.05 ± 0.02	1.00:1
Ddx+Dtx	0.43±0.11	0.39±0.10	1.07:1
Dtx/(Ddx+Dtx)	0.13±0.04	0.13±0.05	1.10:1
Tot Pig	6.19±1.28	5.24±0.74	1.18:1

Table S1.1

Sample to sample variation in photosynthetic pigment content (pg/cell).

Control Sample	Chl c	Fx	Ddx+Dtx	Chl a	β-car
1	0.2453	1.5268	1.4614	5.9333	0.1676
2	0.2457	1.5352	1.4941	6.3008	0.1830
3	0.2335	1.5005	1.4672	6.1047	0.1779
4	0.2450	1.5107	1.4775	5.7806	0.1652
5	0.2346	1.5211	1.4967	5.6772	0.1643
Mean	0.2408	1.5188	1.4794	5.9593	0.1716
SD	0.0062	0.0136	0.0158	0.2500	0.0084
SD as % of Mean	2.57	0.89	1.07	4.20	4.88

Exp. 1 Sample	Chl c	Fx	Ddx+Dtx	Chl a	β-car
1	0.1510	1.0122	0.3793	2.8773	0.0878
2	0.1536	1.0862	0.3960	3.0981	0.0890
3	0.1679	1.0761	0.4025	3.1216	0.0938
4	0.1382	0.9512	0.3516	2.6643	0.0771
5	0.1624	1.1361	0.4252	3.3725	0.0981
Mean	0.1546	1.0523	0.3909	3.0268	0.0892
SD	0.0114	0.0717	0.0274	0.2680	0.0079
SD as % of Mean	7.39	6.81	7.02	8.85	8.83

Exp. 2 Sample	Chl c	Fx	Ddx+Dtx	Chl a	β-car
1	0.1396	0.9989	0.3843	3.6702	0.0760
2	0.1668	1.1247	0.4211	2.8530	0.0908
3	0.1473	1.0888	0.4062	3.7972	0.0816
4	0.1486	1.0872	0.4137	3.1542	0.0914
5	0.1699	1.0391	0.3965	2.6585	0.0709
Mean	0.1544	1.0677	0.4044	3.2266	0.0821
SD	0.0132	0.0491	0.0145	0.4975	0.0090
SD as % of Mean	8.57	4.59	3.57	15.42	10.98

Table S1.2

Nightly biomass loss calculated as the percentage of OD750 gained during the day that was lost at night.

		Undiluted			Diluted							Average ± SD
Exp. 1	Date	10/20	-	-	10/21	10/22	10/23	10/24	10/25	10/26	10/27	-
	% Lost	10.6	-	-	4.7	4.8	8.0	5.6	6.7	5.7	4.7	5.7±1.1
Exp. 2	Date	12/16	12/17	12/18	12/19	12/2	12/21	12/22	12/23	12/24	12/25	-
	% Lost	67.5	32.8	9.0	8.4	6.5	5.2	3.5	3.5	6.7	5.3	5.6±1.6

Table S1.3

Yields at optimal harvesting times vs. 8:00 h.

	Time	Biomass (OD750)	Total TAG (RFU)	Chl a (µg)	Fx (µg)	β-car (µg)	Ddx+Dtx (µg)
Exp. 1	8:00 h	0.10	7.5E8	1.2	0.39	0.03	0.17
	Optimal	0.28	4.4E9	3.3	0.85	0.08	0.29
	Fold difference	2.8	5.9	2.8	2.2	2.7	1.7
Exp. 2	8:00 h	0.18	1.9E9	1.3	0.38	0.03	0.17
	Optimal	0.33	8.8E9	3.1	0.85	0.07	0.28
	Fold difference	1.8	4.6	2.4	2.2	2.3	1.6

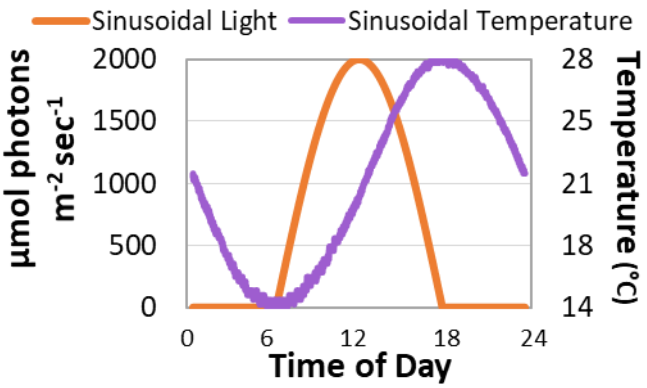


Figure. 1.1.
Sinusoidal changes in light and temperature.

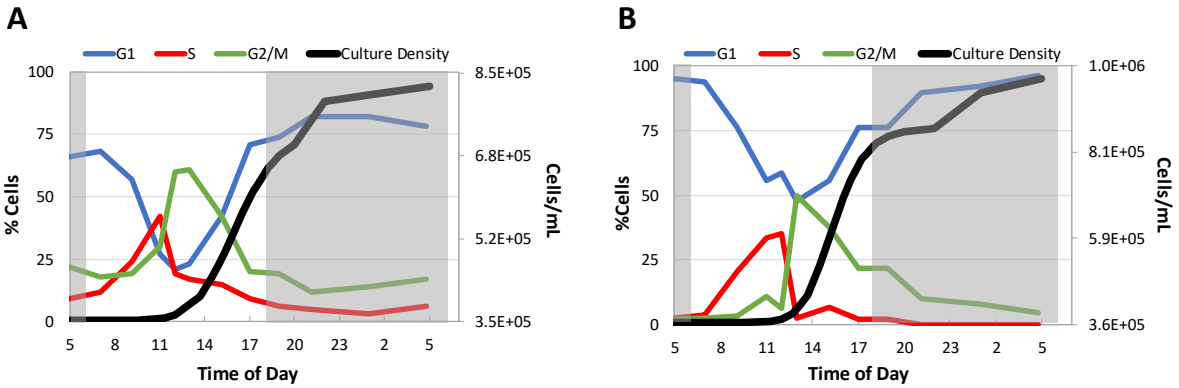


Figure. 1.2.
Culture density relative to cell cycle progression. **A.** Exp. 1, **B.** Exp. 2.

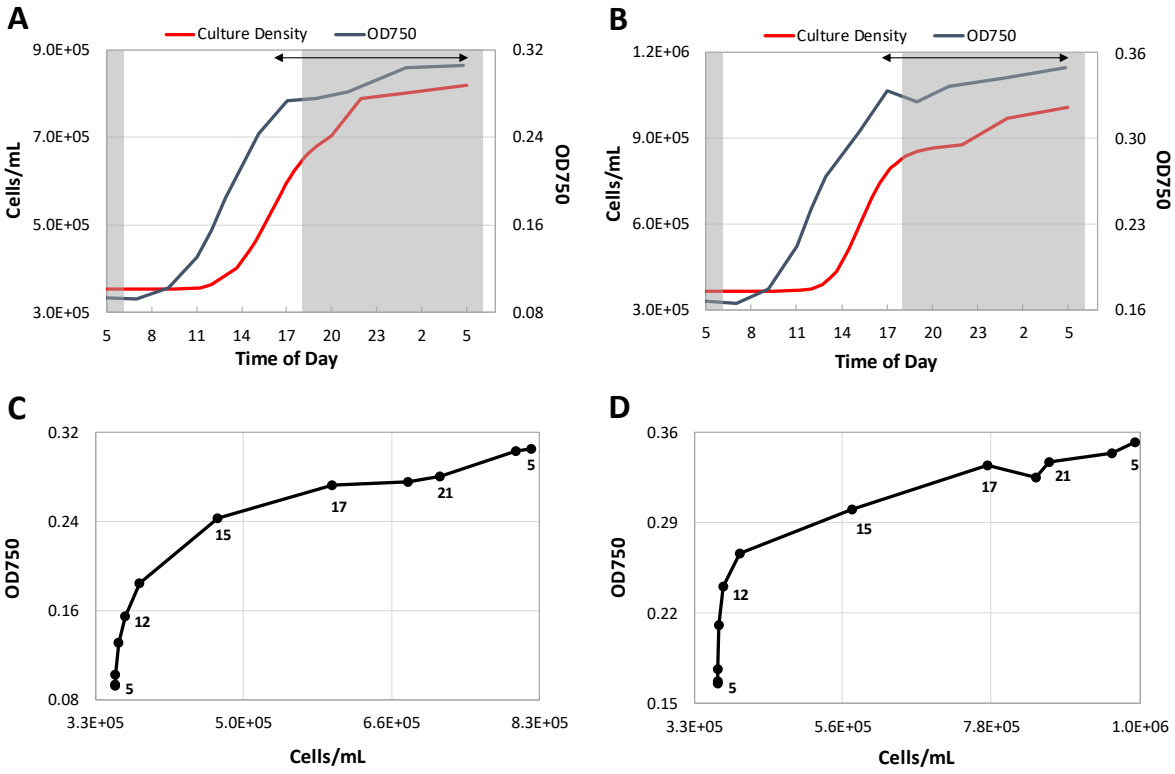


Figure. 1.3.
Culture density and OD750. Arrows indicate optimal biomass harvest times. **A.** Exp. 1 **B.** Exp. 2.; Culture density and corresponding OD750, plotted sequentially.
Numbers on the plot correspond to time of day. **C.** Exp. 1, **D.** Exp. 2.

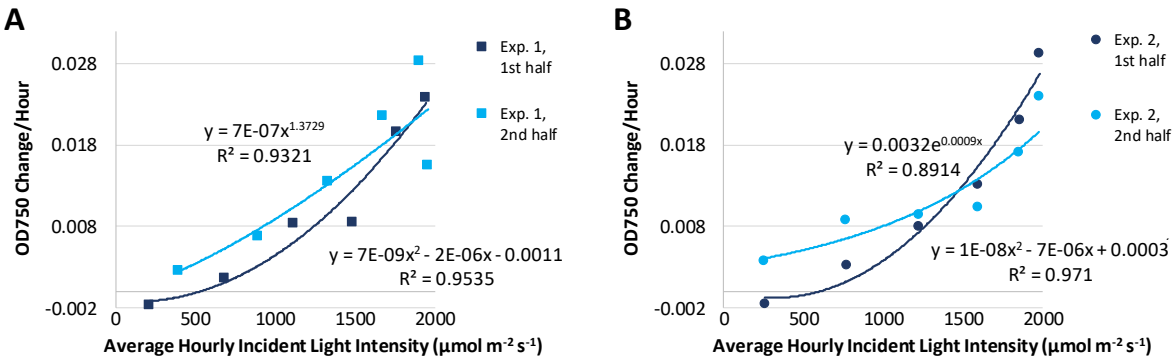


Figure. 1.4.
Rate of OD750 increase per cell versus incident light energy intensity.
First and second halves of the day plotted separately for **A.** Exp. 1, **B.** Exp. 2.

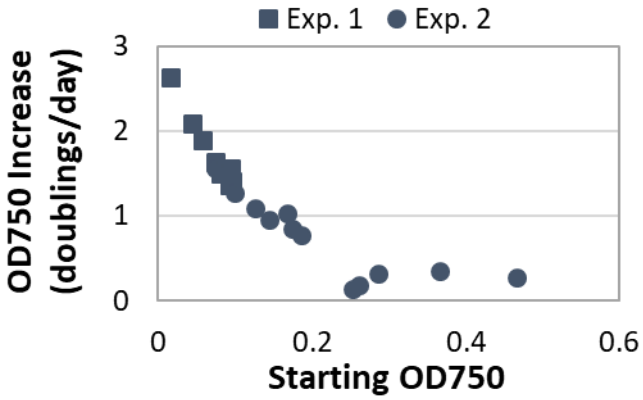


Figure. 1.5.
Increase in OD750 per day versus OD750 at dawn.

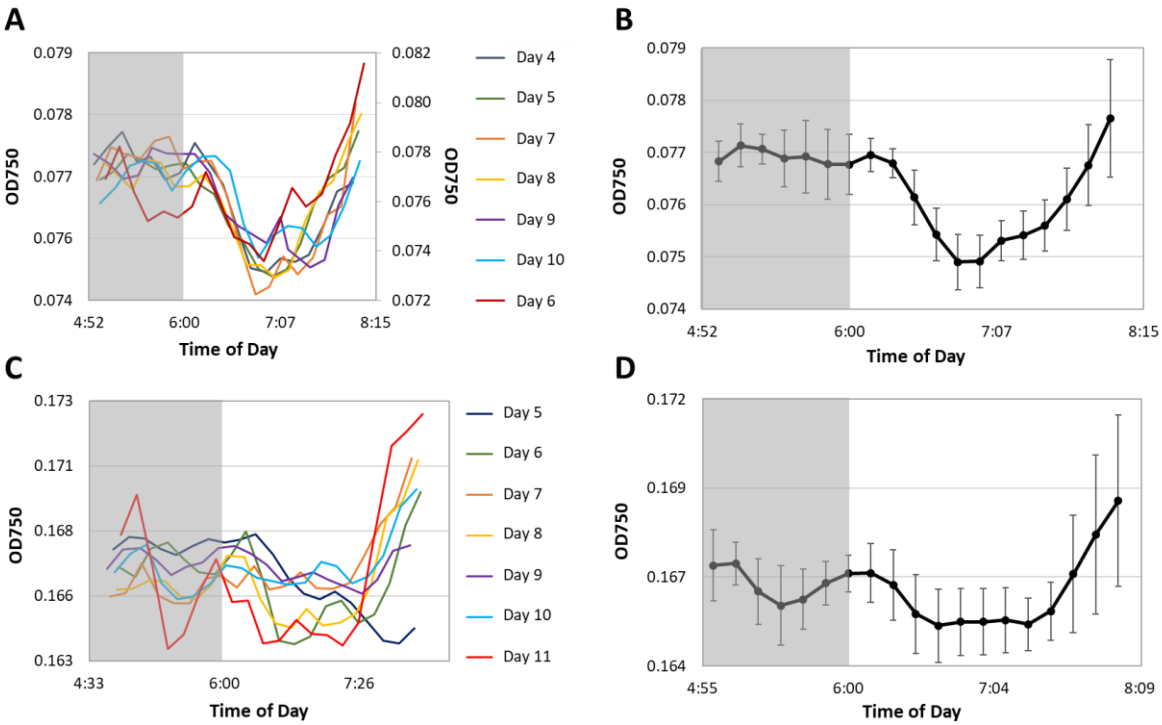


Figure. 1.6.
Dawn dip in OD750.
A. Exp. 1, raw data. Oct. 24 plotted on a secondary y-axis, **B.** Exp. 1, average interpolated values, **C.** Exp. 2, raw data, **D.** Exp. 2, average interpolated value.
Error bars represent 1 standard deviation (SD).

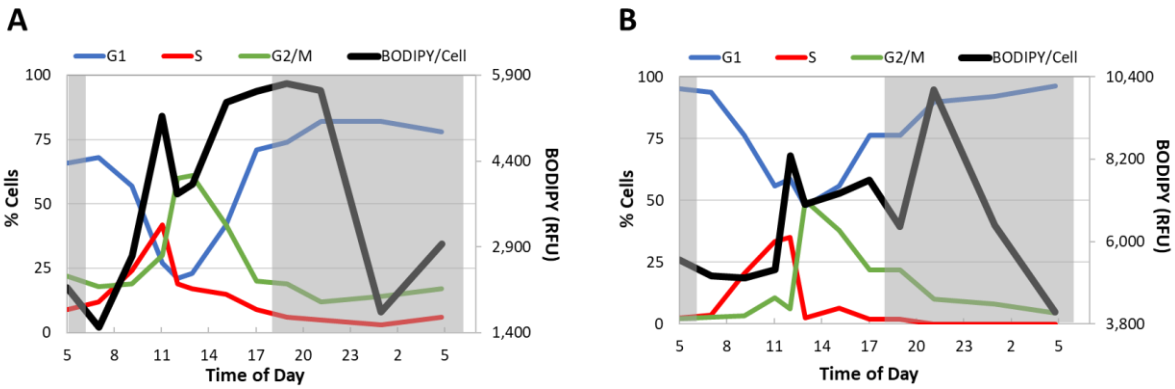


Figure. 1.7.
Cellular TAG levels relative to cell cycle progression **A.** Exp. 1, **B.** Exp. 2.

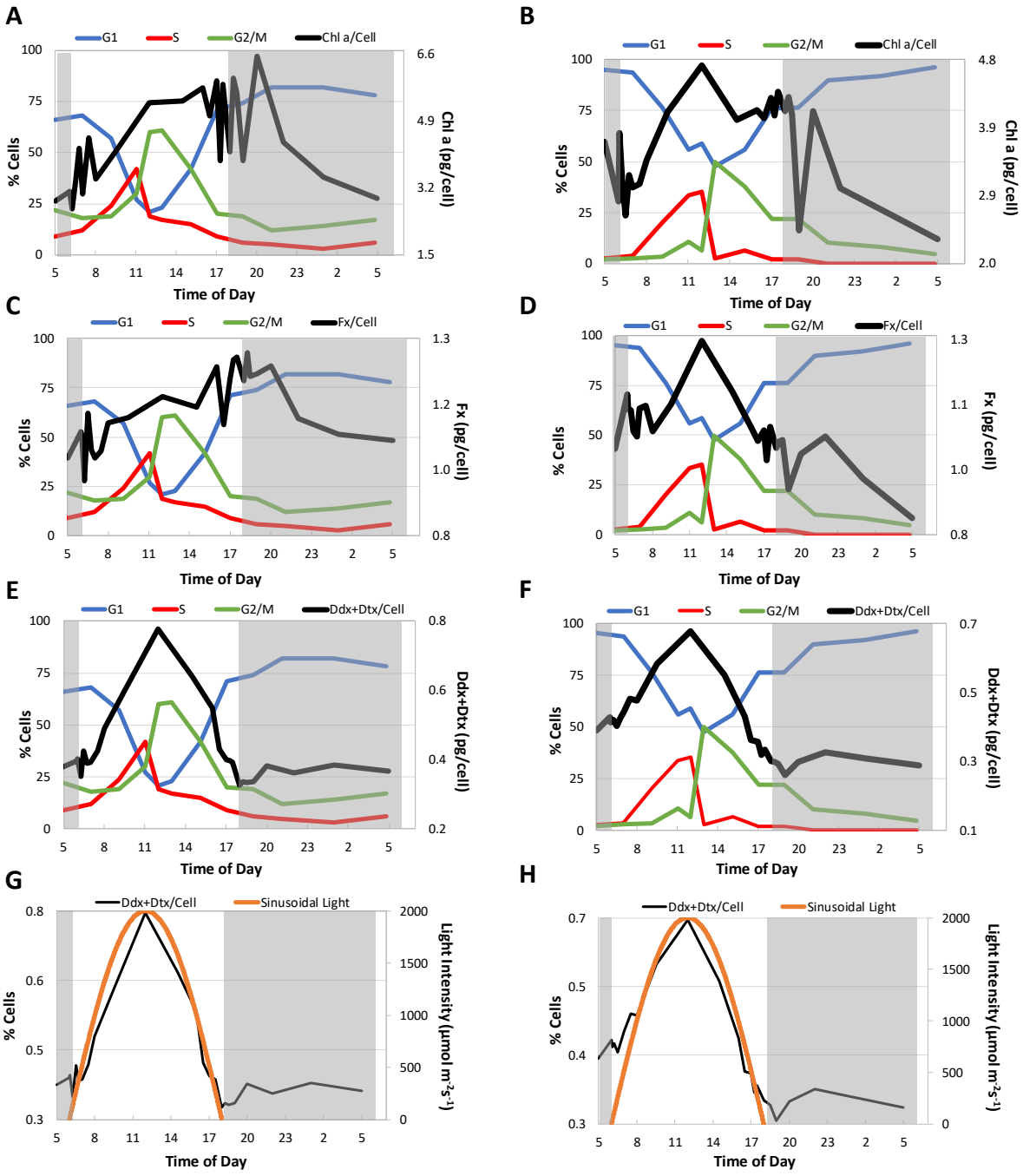


Figure. 1.8.
Changes in cellular pigment abundance vs. cell cycle progression or light intensity. Chl a **A.** Exp.1, **B.** Exp. 2; Fx **C.** Exp.1, **D.** Exp. 2; Ddx+Dtx **(E, G)** Exp. 1, **(F, H)** Exp. 2.

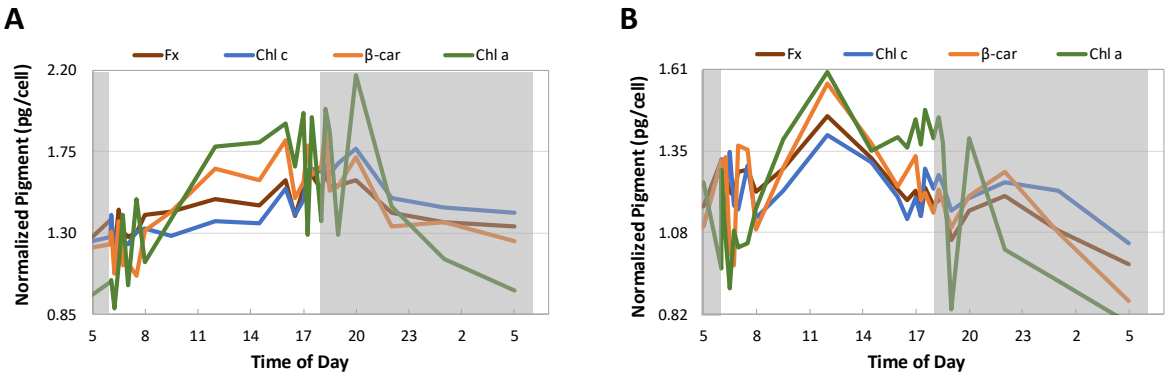


Figure. 1.9.
Changes in cellular abundance of Fx, Chl a, Chl c, and β-car normalized to a common mean.
A. Exp. 1, **B.** Exp. 2

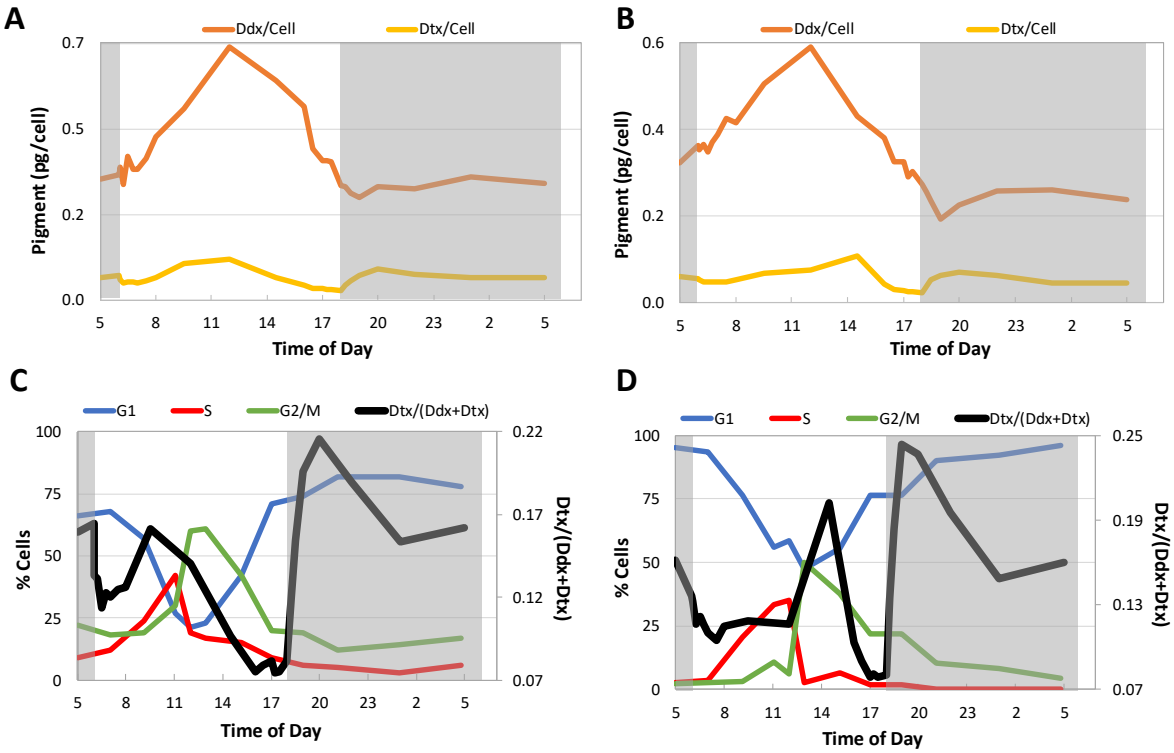


Figure. 1.10.
Ddx and Dtx cellular abundance **A.** Exp. 1, **B.** Exp. 2;
Ddx+Dtx de-epoxidation state, Dtx/(Ddx+Dtx) **C.** Exp. 1, **D.** Exp. 2.

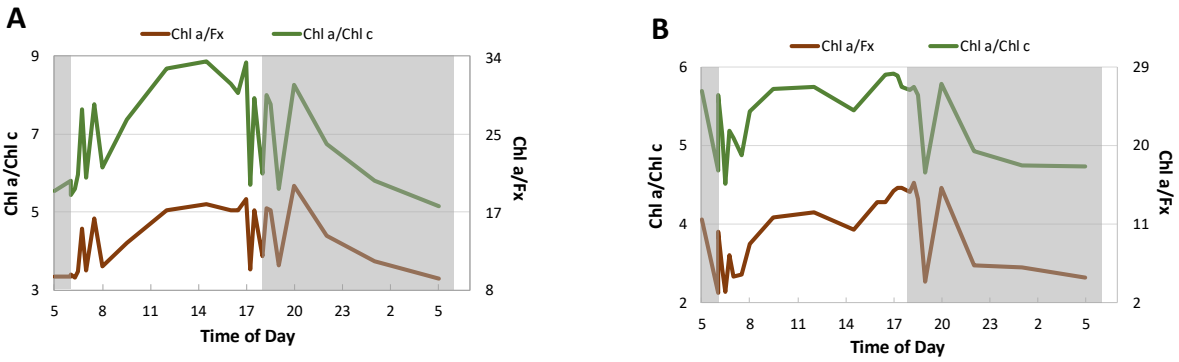


Figure. 1.11.
Changes in the Chl a/Fx and Chl a/Chl c ratios **A.** Exp.1, **B.** Exp. 2.

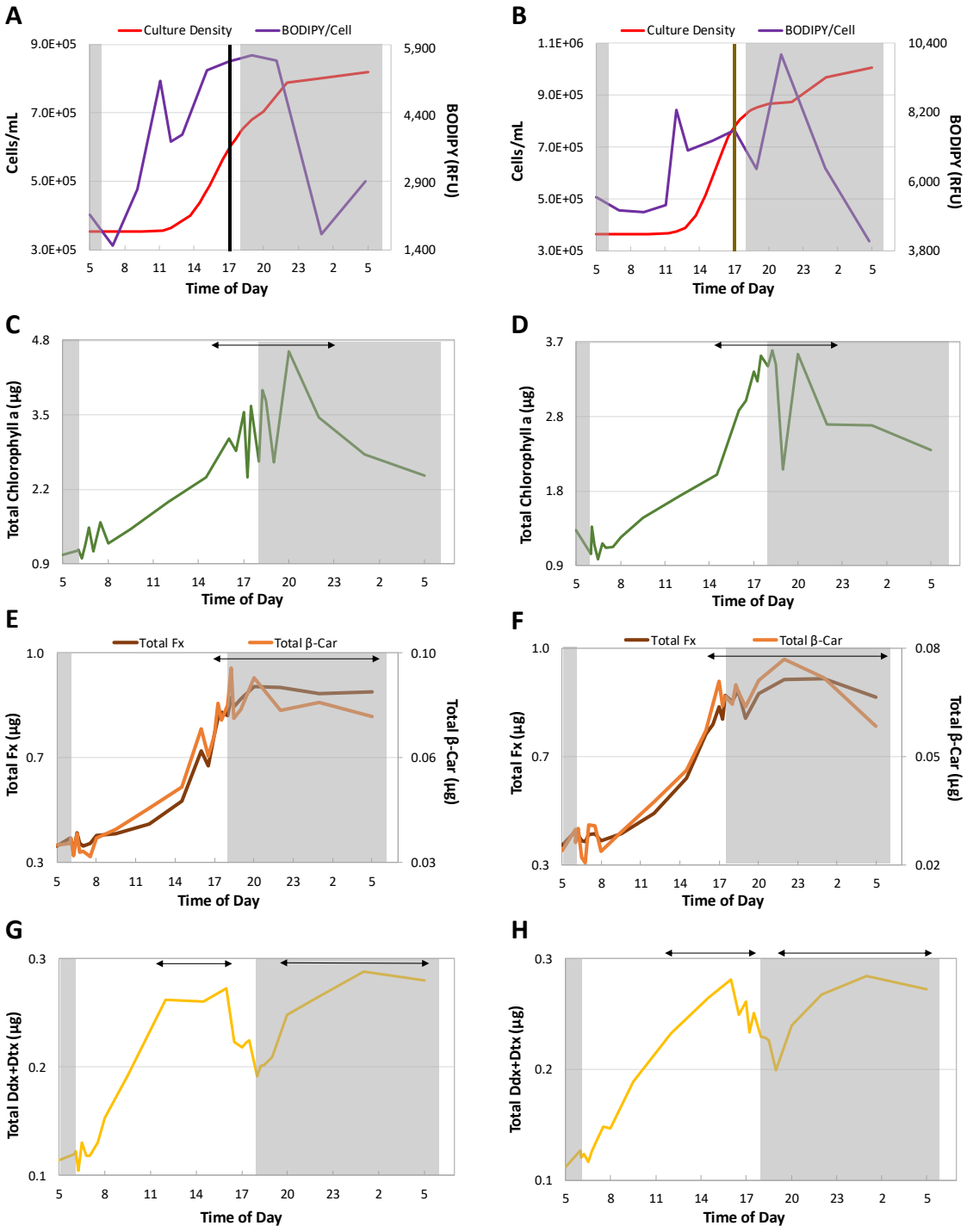


Figure. 1.12.
Optimal harvesting times. Culture density and cellular TAG content.
A. Exp. 1 and **B.** Exp. 2. Black vertical bars indicate optimal harvest times;
Total Chl a **C.** Exp. 1 and **D.** Exp. 2; Total Fx and β-car, **E.** Exp. 1 and **F.** Exp. 2;
Total Ddx+Dtx. **G.** Exp. 1 and **H.** Exp. 2. Black arrows indicate optimal harvest time frames.

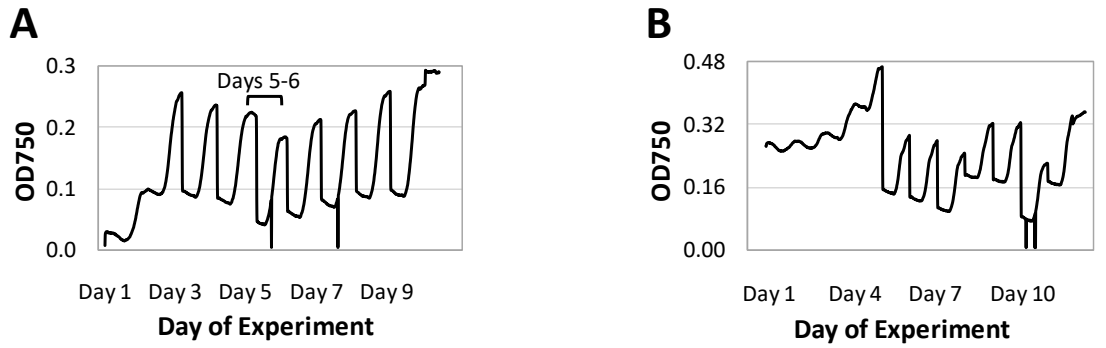


Figure. S1.1.
Daily OD750 changes. **A.** Exp. 1, **B.** Exp. 2.

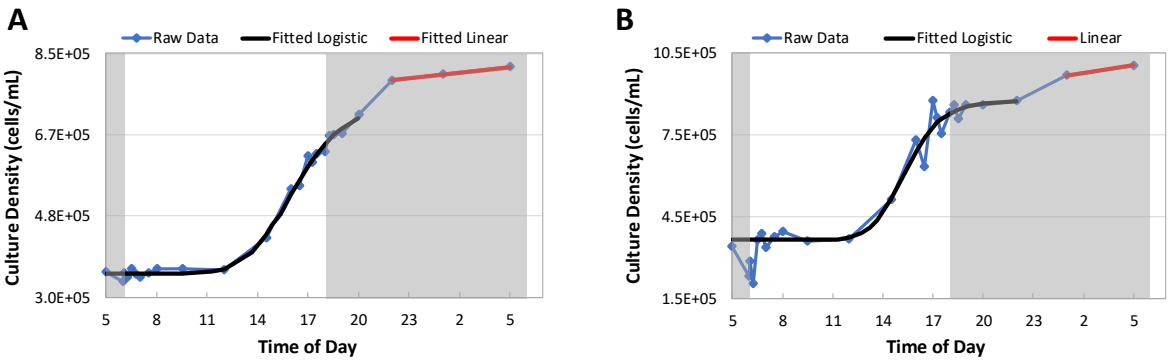


Figure. S1.2.
Raw cell counts and fitted curves.
A. Exp. 1. Fitted Logistic: y = 7.425801 + (3.528496 - 7.425801)/(1 + (x/653.6711)^7.328171), R^2 = 0.99; Fitted Linear: y = 0.0007252252*x + 7.163063, R^2 = 0.99,
B. Exp. 2. Fitted logistic: y = 8.788881 + (3.652112 - 8.788881)/(1 + (x/604.2648)^9.63665, R^2 = 0.98; Linear = y = 0.0015x + 7.8928.

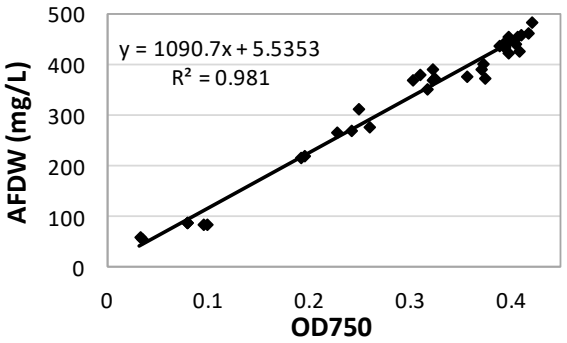


Figure. S1.3.
Correlation between ash-free dry weight (AFDW) and OD750.

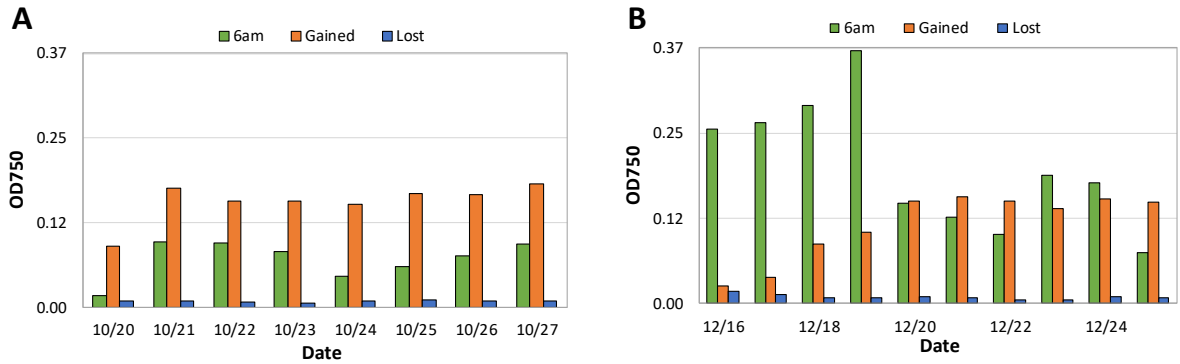


Figure. S1.4.
Biomass at 6am, gained by the end of the light period, and lost at night **A.** Exp. 1, **B.** Exp. 2.

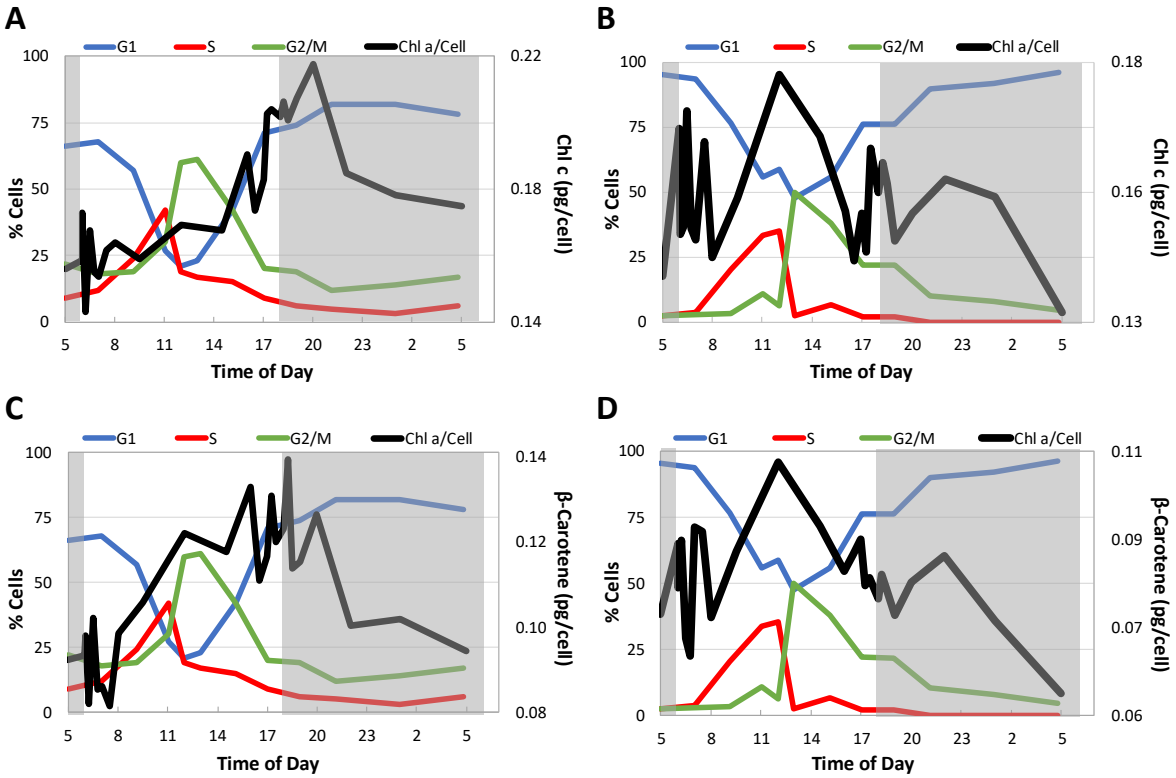


Figure. S1.5.
Changes in pigment abundance vs. cell cycle progression.
Chl c **A.** Exp. 1, **B.** Exp. 2; β-car **C.** Exp. 1, **D.** Exp. 2.

Figure. S1.6.
Total TAG (combined TAG content of all the cells in the culture) versus cell cycle progression.
A. Exp. 1, **B.** Exp. 2.

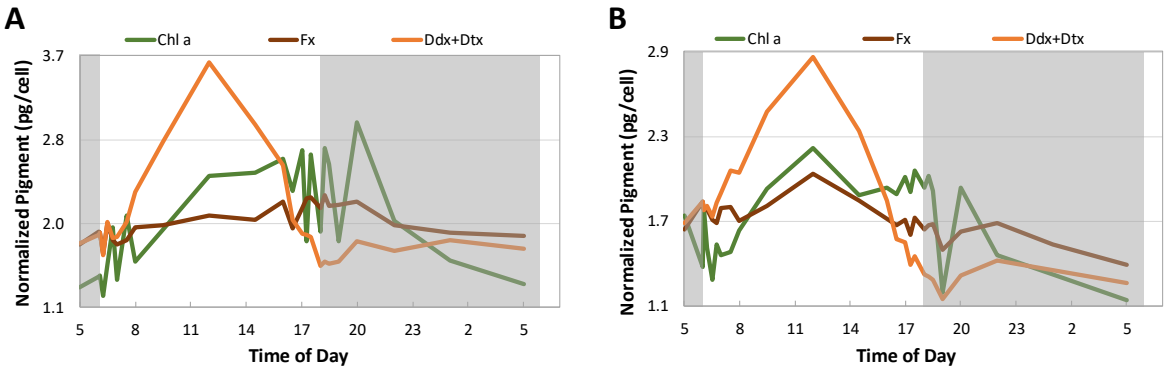


Figure. S1.7.
Changes in cellular Chl a, Fx, and Ddx+Dtx abundance, normalized to a common mean.
A. Exp. 1, **B.** Exp. 2.

1.8 REFERENCES

R. Abbriano, N. Vardar, D. Yee, M. Hildebrand, Manipulation of a glycolytic regulator alters growth and carbon partitioning in the marine diatom *Thalassiosira pseudonana*. Algal Res. 32 (2018) 250-258.

L. W. J. Anderson, B. M. Sweeney, Diel changes in sedimentation characteristics of *Ditylum brightwellii*: changes in cellular lipid and effects of respiratory inhibitors and ion-transport modifiers. Limnol Oceanogr. 22 (3) (1977) 539-552.

K. Athenstaedt, G. Daum, The life cycle of neutral lipids: synthesis, storage and degradation. Cel Mol Life Sci. 63 (12) (2006) 1355-1369.

A. I. Barros, A. L. Goncalves, M. Simoes, J. C. M. Pires, Harvesting techniques applied to microalgae: a review. Renew Sust Energ Rev. 41 (2015) 1489-1500.

H. Bonnefond, N. Moelants, A. Talec, O. Bernard, A. Sciandra, Concomittant effects of light and temperature diel variations on the growth rate and lipid production of *Dunaliella salina*. Algal Res. 14 (2016) 72-78.

M. S. Bono Jr., R. D. Garcia, D. V. Sri-Jayantha, B. A. Ahner, B. J. Kirby, Measurement of lipid accumulation in *Chlorella vulgaris* via flow cytometry and liquid-state ^{1}H NMR spectroscopy for development of an NMR-traceable flow cytometry protocol. PLoS One 10 (8) (2015) e0134846.

C. Buchel, Fucoxanthin-chlorophyll proteins in diatoms: 18 and 19 kDa subunits assemble into different oligomeric states. Biochem. 42 (2003) 13027-13034.

J. E. Burris, Photosynthesis, photorespiration, and dark respiration in eight species of algae. Mar Biol. 39 (1977) 371-379.

M. S. Chauton, P. Winge, T. Brembu, O. Vadstein, A. M. Bones, Gene regulation of carbon fixation, storage, and utilization in the diatom *Phaeodactylum tricornutum* acclimated to light/dark cycles. Plant Physiol. 161 (2) (2013) 1034-1048.

M. Chioccioli, B. Hankamer, I. L. Ross, Flow cytometry pulse with data enables rapid and sensitive estimation of biomass dry weight in the microalgae *Chlamydomonas reinhardtii* and *Chlorella vulgaris*. PLoS One 9 (5) (2014) e97269.

S. W. Chisholm, L. E. Brand, Persistence of cell division phasing in marine phytoplankton in continuous light after entrainment to light:dark cycles. J Exp Mar Biol Ecol. 51 (1981) 107-118.

S. W. Chisholm, F. M. M. Morel, W. S. Slocum, The phasing and distribution of cell division cycles in marine diatoms, in: P. G. Falkowski (Ed.), Primary productivity in the sea, Environmental Science Research Vol. 19, Springer, Boston MA, 1980.

D. R. Clark, K. J. Flynn, N. J. P. Owens, The large capacity for dark nitrate-assimilation in diatoms may overcome nitrate limitation on growth. New Phytol. 155 (1) (2002) 101-108.

R. Croce, H. van Amerongen, Natural strategies for photosynthetic light harvesting. Nat Chem Biol. 10 (2014) 492-501.

R. L. Cuhel, D. R. S. Lean, Influence of light intensity, light quality, temperature, and daylength on uptake and assimilation of carbon dioxide and sulfate by lake plankton. Can J Fish Aquat Sci. 44 (1987) 2118-2132.

R. L. Cuhel, P. B. Ortner, D. R. S. Lean, Night synthesis of protein by algae. Limnol Oceanogr. 29 (4) (1984) 731-744.

W. Darley, B. Volcani, Role of silicon in diatom metabolism, A silicon requirement for deoxyribonucleic acid synthesis in the diatom *Cylindrotheca fusiformis* Reimann and Lewin. Exp. Cell. Res. 58 (1969) 334–42.

W. M. Darley, B. E. Volcani, Synchronized cultures: diatoms. Methods Enzymol. 23 (1971) 85-96.

V. Ebenzer, L. K. Medlin, J. Ki, Molecular detection, quantification, and diversity evaluation of microalgae. Mar Biotechnol. 14 (2012) 129-142.

S. J. Edmundson, M. H. Huesemann, The dark side of algae cultivation: characterizing night biomass loss in three photosynthetic algae, *Chlorella sorokiniana*, *Nannochloropsis salina* and *Picochlorum* sp. Algal Res. 12 (2015) 470-476.

R. W. Eppley, R. W. Holmes, E. Paasche, Periodicity in cell division and physiological behavior of *Ditylum brightwelii*, a marine planktonic diatom, during growth in light-dark cycles. Arch Mikrobiol. 56 (1967) 305-323.

P. G. Falkowski, J. A. Raven, Aquatic photosynthesis (second edition). Princeton University Press, 2007.

M. W. Fawley, Effects of light intensity and temperature interactions on growth characteristics of *Phaeodactylum tricornutum* (Bacillariophyceae). J Phycol. 20 (1984) 67-72.

R. J. Geider, B. A. Osborne, Light absorption by a marine diatom: experimental observations and theoretical calculations of the package effect in a small *Thalassiosira* species. Mar Biol. 96 (1987) 299-308.

R. J. Geider, B. A. Osborne, Respiration and microalgal growth: a review of the quantitative relationship between dark respiration and growth. New Phytol. 112 (1989) 327-341.

M. L. Gerardo, S. V. D. Hende, H. Vervaeren, T. Coward, S. C. Skill, Harvesting of microalgae within a biorefinery approach: a review of the developments and case studies from pilot-plants. Algal Res. 11 (2015) 248-262.

J. M. Goodman, The gregarious lipid droplet J Biol Chem. 283 (42) (2008) 28005-28009.

R. Goss, E. A. Pinto, C. Wilhelm, M. Richter, The importance of highly active and ΔpH-regulated diatoxanthin epoxidase for the regulation of the PSII antenna function in diadinoxanthin cycle containing algae. J Plant Physiol. 163 (10) (2006) 1008-1021.

E. M. Grima, J. A. S. Perez, F. G. Camacho, J. M. F. Sevilla, F. G. A. Fernandez, J. U. Cardona, Biomass and icosapentoic acid productivities from an outdoor batch culture of *Phaeodactylum tricornutum* UTEX 640 in an airlift tubular photobioreactor. Appl Microbiol Biotechnol. 42 (1995) 658-663.

J. U. Grobbelaar, C. J. Soeder, Respiration losses in planktonic green algae cultivated in raceway ponds. J Plankton Res. 7 (4) (1985) 497-506.

I. Grouneva, T. Jakob, C. Wilhelm, R. Goss, The regulation of xanthophyll cycle activity and of non-photochemical fluorescence quenching by two alternate electron flows in the diatoms *Phaeodactylum tricornutum* and *Cyclotella meneghiniana*. Biochim Biophys Acta. 1787 (2009) 929-938.

H. Guterman, S. Ben-Yakov, A. Vonshak, Automatic on-line growth estimation method for outdoor algal biomass production. Biotechnol Bioeng. 34 (2) (1988) 143-152.

S. Ho, C. Chen, J. Chang, Effect of light intensity and nitrogen starvation on CO_2 and lipid/carbohydrate production of an indigenous microalga *Scenedesmus obliquus* CNW-N. Bioresour Technol. 113 (2012) 244-252.

H. Hoogenhout, Synchronous cultures of algae. Phycologia 2 (4) (1963) 135-147.

Q. Hu, N. Kurano, M. Kawachi, I. Iwasaki, S. Miyachi, Ultrahigh-cell-density culture of a marine green alga *Chlorococcum littorale* in a flat-plate photobioreactor. Appl Microbiol Biotechnol. 49 (1998) 655-662.

Q. Hu, M. Sommerfield, E. Jarvis, M. Ghirardi, M. Posewitz, M. Seibert, A. Darzins, Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. Plant J. 54 (4) (2008) 621-639.

M. Huesemann, T. Dale, A. Chavis, B. Crowe, S. Twary, A. Barry, D. Valentine, R. Yoshida, M. Wigmosta, V. Cullinan, Simulation of outdoor pond cultures using indoor LED-lighted and temperature-controlled raceway ponds and Phenometrics photobioreactors. Algal Res. 21 (2017) 178-190.

M. Huesemann, A. Chavis, S. Edmundson, D. Rye, M. Wigmosta, Climate-simulated pond culturing: quantifying the maximum achievable annual biomass productivity of *Chlorella sorokiniana* (DOE 1412) in the United States, J Appl Phycol. 30 (2018) 287-298.

G. F. Humphrey, Photosynthetic characteristics of algae grown under constant illumination and light-dark regimes. J Exp Mar Biol Ecol. 40 (1979) 63-70.

T. Jakob, R. Goss, C. Wilhelm, Activation of diadinoxanthin de-epoxidase due to a chlororespiratory proton gradient in the dark in the diatom *Phaeodactylum tricornutum*. Plant Biol. 1 (1999) 76-82.

T. Jakob, R. Goss, C. Wilhelm, Unusual pH-dependence of diadinoxanthin de-epoxidase activation causes chlororespiratory induced accumulation of diatoxanthin in the diatom *Phaeodactylum tricornutum*. J Plant Physiol. 158 (2001) 383-390.

D. Jallet, M. A. Caballero, A. A. Gallina, M. Youngblood, G. Peers, Photosynthetic physiology and biomass partitioning in the model diatom *Phaeodactylum tricornutum* grown in a sinusoidal light regime. Algal Res. 18 (2016) 51-60.

L. Joseph, T. A. Villareal, Nitrate reductase activity as a measure of nitrogen incorporation in *Rhizosolenia formosa* (H. Peragallo): internal nitrate and diel effects. J Exp Mar Biol Ecol. 229 (1998) 159-176.

S. Khanra, M. Mondal, G. Halder, O. N. Tiwari, K. Gayen, T. K. Bhowmick, Downstream processing of microalgae for pigments, protein and carbohydrate in industrial application: a review. Food Bioprod Process. (2018) https://doi.org/10.1016/j.fbp.2018.02.002

K. Kohata, M. Watanabe, Diel changes in the composition of photosynthetic pigments and cellular carbon and nitrogen in *Chantonella antiqua* (Raphidophyceae). J Phycol. 24 (1988) 58-66.

K. Kohata, M. Watanabe, Diel changes in the composition of pigments and cellular carbon and nitrogen in *Pyramimonas parkeae* (Prasinophyceae). J Phycol. 25 (1989) 377-385.

W.A. Kozlowski, D. Deutschman, I. Garibotti, C. Trees, M. Vernet, An evaluation of the application of CHEMTAX to Antarctic coastal pigment data. Deep-Sea Res. Pt I 58 (2011) 350-364.
P. Kuczynska, M. Jemiola-Rzeminska, K. Strzalka, Photosynthetic pigments in diatoms. Mar Drugs 13 (9) (2015) 5847-5881.

C. F. Kurat, H. Wolinski, J. Petschnigg, S. Kaluarachchi, B. Andrews, K. Natter, S. D. Kohlwein, Cdk1/Cdc28-dependent activation of the major triacylglycerol lipase Tgl4 in yeast links lipolysis to cell-cycle progression. Mol Cell 33 (1) (2009) 53-63.

T. Lacour, A. Sciendra, A. Talec, P, Mayzaud, O. Bernard, Diel variations of carbohydrates and neutral lipids in nitrogen-sufficient and nitrogen-starved cyclostat cultures of *Isochrysis* sp. J Phycol. 48 (2012) 966-975.

C. Lancelot, S. Mathot, Biochemical fractionation of primary production by phytoplankton in Belgian coastal waters during short- and long-term incubations with ^{14}C-bicarbonate. Mar Biol. 86 (1985) 219-226.

F. Le Borgne, J. Pruvost, Investigation and modeling of biomass decay rate in the dark and its potential influence on netproductivity of solar photobioreactors for microalga *Chlamydomonas reinhardtii* and cyanobacterium *Arthrospira platensis*. Bioresour. Technol. 138 (2013) 271-276.

D. R. S. Lean, R. L. Cuhel, M. N. Charlton, Protein synthesis: a measure of growth for lake plankton. Hydrobiologia 173 (1989) 119-126.

B. Lepetit, R. Goss, T. Jakob, C. Wilhelm, Molecular dynamics of the diatom thylakoid membrane under different light conditions. Photosynth Res. 111 (2012) 245-257.

B. F. Lucker, C. C. Hall, R. Zegarac, D. M. Kramer, The environmental photobioreactor (ePBR): an algal culturing platform for simulating dynamic natural environments. Algal Res. 6 (2014) 242-249.

Y. Maeda, D. Nojima, T. Yoshino, T. Tanaka, Structure and properties of oil bodies in diatoms. Phil Trans R Soc B. 372 (2017) 20160408.

J. Matos, C. Cardoso, N. M. Bandarra, C. Afonso, Microalgae as healthy ingredients for functional food: a review. Food Funct. 8 (2017) 2672-2685.

M. H. A. Michels, P. M. Slegers, M. H. Vermue, R. H. Wijffels, Effect of biomass concentration on the productivity of *Tetraselmis suecica* in a pilot-scale tubular photobioreactor using natural sunlight. Algal Res. 4 (2014) 12-18.

S. Miyagishima, Regulation of cell cycle progression by circadian rhythms in *Cyanidioschyzon merolae*, in: T. Kuroiwa, S. Miyagishima, S. Matsunaga, N. Sato, H. Nozaki, K. Tanaka, O. Misumi (Eds.), *Cyanidioschyzon merolae*: a new model eukaryote for cell and organelle biology, Springer, Singapore, 2017.

I. Morris, W. Skea, Products of photosynthesis in natural populations of marine phytoplankton from the Gulf of Maine. Mar Biol. 47 (1978) 303-312.

S. M. Myklestad, Production, chemical structure, metabolism, and biological function of the (1-3)-linked, β3-D-glucans in diatoms. Biol Oceanogr. 6 (1989) 313-326.

D. M. Nelson, L. E. Brand, Cell division periodicity in 13 species of marine phytoplankton on a light:dark cycle. J Phycol. 15 (1979) 67-75.

M. Nymark, K. C. Valle, T. Brembu, K. Hancke, P. Winge, K. Andersen, G. Johnsen, A. M. Bones, An integrated analysis of molecular acclimation to high light in the marine diatom *Phaeodactylum tricornutum*. PLoS One. 4 (11) (2009) e7743.

J. C. Ogbonna, H. Tanaka, Night biomass loss and changes in biochemical composition of cells during light/dark cyclic culture of *Chlorella pyrenoidosa*. J Ferment Bioeng. 82 (6) (1996) 558-564.

I. Orefice, R. Chandrasekaran, A. Smerilli, F. Corato, T. Caruso, A. Casillo, M. M. Corsaro, F. D. Piaz, A. V. Ruban, C. Brunet, Light-induced changes in the photosynthetic physiology and biochemistry in the diatom *Skeletonema marinoi*. Algal Res. 17 (2016) 1-13.

T. G. Owens, P. G. Falkowski, T. E. Whitledge, Diel periodicity in cellular chlorophyll content in marine diatoms. Mar Biol. 59 (1980) 71-77.

E. Paasche, Marine plankton algae grown with light-dark cycles. II. *Ditylum brightwelii* and *Nitzschia turgidula*. Physiol Plant. 21 (1968) 66-77.

G. Peltier, J. Ravenel, A. Vermeglio, Inhibition of a respiratory activity by short saturating flashes in *Chlamydomonas*: evidence for a chlororespiration. Biochim Biophis Acta. 893 (1987) 83-90.

A. F. Post, Z. Dubinsky, K. Wyman, P. G. Falkowski, Kinetics of light-intensity adaptation in a marine phytoplankton. Mar Biol. 83 (1984) 231-238.

M. Ragni, M. R. d'Alcala, Circadian variability in the photobiology of *Phaeodactylum tricornutum*: pigment content. J Plankton Res. 29 (2) (2007) 141-156.

P. Raimbault, M. Mingazzini, Diurnal variations of intracellular nitrate storage by diatoms: effects of nutritional state. J Exp Mar Biol Ecol. 112 (1987) 217-232.

J. A. Raven, R. J. Geider, Temperature and algal growth. New Phytol. 110 (1988) 441-461.

R. B. Rivkin, Carbon-14 labelling patterns of individual marine phytoplankton from natural populations. Mar Biol. 89 (1985) 135-142.

P. G. Roessler, Environmental control of glycerolipid metabolism in microalgae: commercial implications and future research directions. J Phycol. 26 (1990) 393-399.

R. E. H. Smith, P. Clement, E. J. Head, Night metabolism of recent photosynthate by sea ice algae in the high Arctic. Mar Biol. 107 (1990) 255-261.

A. E. Solovchenko, Physiological role of neutral lipid accumulation in eukaryotic microalgae under stresses. Russ J Plant Physiol. 59 (2) (2012) 167-176.

A. Sukenik, J. Bennett, P. Falkowski, Light-saturated photosynthesis – limitation by electron transport or carbon fixation. Biochim Biophis Acta. 891 (1987) 205-215.

A. Sukenik, Y. Carmeli, Lipid synthesis and fatty acid composition in *Nannochloropsis* sp. (Eustigmatophyceae) grown in a light-dark cycle. J Phycol. 26 (1990) 463-469.

Y. Sumi, Microalgae pioneering the future-application and utilization. Life Sci Res Unit Quart Rev. 34 (2009) 9-21.

B. Tamburic, S. Guruprasad, D. T. Radford, M. Szabo, R. M. Lilley, A. W. D. Larkum, J. B. Franklin, D. M. Kramer, S. I. Blackburn, J. A. Raven, M. Schliep, P. J. Ralph, The effect of diel temperature and light cycles on the growth of *Nannochloropsis oculata* in a photobioreactor matrix. PLoS One 9 (1) (2014) e86047.

G. A. Thompson, Lipids and membrane function in green algae. Biochim Biophis Acta. 1302 (1996) 17-45.

G. Torzillo, A. Sacchi, P. Materassi, A. Richmond, Effect of temperature on yield and night biomass loss in *Spirulina platensis* grown outdoors in tubular photobioreactors. J Appl Phycol. 3 (1991) 103-109.

J. C. Traller, M. Hildebrand, High throughput imaging to the diatom *Cyclotella cryptica* demonstrates substantial cell-to-cell variability in the rate and extent of triacylglycerol accumulation. Algal Res. 2 (2013) 244-252.

J. C. Traller, S. J. Cokus, D. A. Lopez, O. Gaidarenko, S. R. Smith, J. P. McCrow, S. D. Gallaher, S. Podell, M. Thompson, O. Cook, M. Morselli, A. Jaroszewicz, E. E. Allen, A. E. Allen, S. S. Merchant, M. Pellegrini, M. Hildebrand, Genome and methylome of the oleaginous diatom *Cyclotella cryptica* reveal genetic flexibility toward a high lipid phenotype. Biotechnol Biofuels. 9 (2016) 258.

K. M. Varum, S. Mylkestad, Effects of light, salinity, and nutrient limitation on the production of β-1,3-D-glucan and exo-D-glucanase activity in *Skeletonema costatum* (Grev.) Cleve. J Exp Mar Biol Ecol. 83 (1984) 13-25.

K. M. Varum, K. Ostgaard, K. Grimsrud, Diurnal rhythms in carbohydrate metabolism of the marine diatom *Skeletonema costatum* (Grev.) Cleve. J Exp Mar Biol Ecol. 102 (1986) 249-256.

C. E. Williamson, Phased cell division in natural and laboratory populations of marine planktonic diatoms. J Exp Mar Biol Ecol. 43 (1980) 271-279.

W. Yang, S. Zou, M. He, C. Fei, W. Luo, S. Zheng, B. Chen, C. Wang, Growth and lipid accumulation in three *Chlorella* strains from different regions in response to diurnal temperature fluctuations. Bioresour Technol. 202 (2016) 15-24.

C. S. Yentsch, R. F. Scagel, Diurnal study of phytoplankton pigments: an *in situ* study in East Sound, Washington. J Mar Res. 17 (1958) 567-581.

J. K. Zehmer, Y. Huang, G. Peng, J. Pu, R. G. W. Anderson, P. Liu, A role for lipid droplets in inter-membrane lipid traffic. Proteomics. 9 (2009) 914-921.

CHAPTER 2

Novel *Thalassiosira pseudonana* violaxanthin de-epoxidase-like enzyme (VDL2) catalyzes fucoxanthin biosynthesis

2.1 ABSTRACT

Despite the ubiquity and ecological importance of diatoms, much remains to be understood about their physiology and metabolism, including their carotenoid biosynthesis pathway. Early carotenoid biosynthesis steps are well-conserved. The sequence of, and enzymes that catalyze, the later steps that lead to the major accessory photopigment fucoxanthin (Fx) and main photoprotective pigment pool comprised of diadinoxanthin (Ddx) and its reversibly de-epoxidized form diatoxanthin (Dtx), remain unclear. We used sequence comparison to known carotenoid biosynthesis enzymes to identify novel candidates in the diatom *Thalassiosira pseudonana*. RNA-seq data was used to create full-length gene models, and we focused on those that encode proteins predicted to be chloroplast-localized. Based on differential transcriptional regulation of the first two biosynthetic steps, we propose that there are two ways to independently initiate carotenoid biosynthesis. Chloroplast division, which requires the population of newly divided chloroplasts with photopigments, is one proposed carotenogenic stimulus. Irradiance increase, which triggers accumulation of the Ddx+Dtx, is the other. Based on transcriptomic and physiological data, we also propose that the two predicted *T. pseudonana* zeaxanthin epoxidases have distinct roles. We identified a novel violaxanthin de-epoxidase-like enzyme (Thaps3_11707, VDL2) that when overexpressed increases Fx abundance while stoichiometrically reducing Ddx+Dtx. Based on transcriptomics, we propose that Thaps3_10233 may also contribute to Fx biosynthesis. Unlike other carotenoid biosynthesis genes, VDL2 and Thaps3_10233 appear transcriptionally co-regulated with photoantenna proteins, suggesting a mechanism of diverting some precursors towards Fx rather than Ddx+Dtx when necessary to populate photoantenna proteins. Separately using antisense to target VDL2, VDL1, and both LUT1-like copies simultaneously, reduced the overall cellular photopigment content, including chlorophylls, suggesting a destabilization of light-harvesting complexes by Fx deficiency.

2.2 INTRODUCTION

Diatoms are incredibly diverse, environmentally flexible, and productive photosynthetic microalgae that belong to the Stramenopile or heterokont class. There are over 100,000 estimated diatom species that occupy a wide variety of habitats. Diatoms are of great ecological importance, as they are responsible for approximately 40% of primary marine productivity and play a major role in the global cycling of carbon, nitrogen, phosphorus, and silicon [Hildebrand et al. 2012, Wilhelm et al. 2006]. They have a complex evolutionary history that includes a secondary endosymbiotic event, wherein an ancient red alga was engulfed by a heterotrophic eukaryote, at least 800Ma ago. Because of this event, diatoms possess physiological and metabolic features that differ from other algal groups, such as the more extensively studied chlorophytes [Hildebrand et al. 2012, Wilhelm et al. 2006].

Photosynthetic pigments are among the distinguishing features of diatoms. Carotenoids are utilized by all known photosynthetic organisms, and the initial part of the carotenoid biosynthesis pathway from phytoene to β-carotene (β-car) is well-conserved [Bertrand 2010, Coesel et al. 2008]. β-car is synthesized from lycopene, as is α-carotene (α-car) in many organisms. Both β- and α-car have a variety of derivatives that can differ substantially depending on the taxonomic class, sometimes with further inter-species differences. Diatoms do not possess the α-car branch of carotenoids, and their main photoprotective and accessory light-harvesting pigments are β-car derivatives. Knowledge about the sequence of the post-β-car biosynthetic steps in diatoms, as well as the enzymes that catalyze them, is still very sparse [Bertrand 2010, Coesel et al. 2008]. One of the final products of the pathway **(Fig. 2.1)** is fucoxanthin (Fx), the main accessory light-harvesting pigment in diatoms that is responsible for their characteristic golden-brown color. It is bound to photoantenna proteins along with chlorophylls a and c (Chl a, Chl c). Most diatoms have Chl c1 and Chl c2, although Chl c3 has been found in some species as well. Unlike chlorophytes, diatoms do not

make chlorophyll b [Wilhelm et al. 2006]. Diadinoxanthin (Ddx) and diatoxanthin (Dtx) are the other final products of diatom carotenoid biosynthesis. They form a xanthophyll cycle by reversibly interconverting via epoxidation and de-epoxidation **(Fig. 2.1)**, comprising the major photoprotective mechanism employed by diatoms [Goss et al. 2006, Lohr and Wilhelm 1999, Wilhelm et al. 2006]. Some Ddx molecules are bound by photoantenna proteins, while others are dissolved in the lipid shield that surrounds photoantennae [Lepetit et al. 2010]. The other xanthophyll cycle found in diatoms is the violaxanthin (Vx) cycle, in which zeaxanthin (Zx) is reversibly converted to (Vx) via antheraxanthin (Ax) by two epoxidation steps. Both cycles serve to allow the switch between light-harvesting and light-dissipative states as a rapid adjustment to irradiance changes, with the epoxide forms corresponding to the former and de-epoxides to the latter. The Vx cycle is conserved and plays a major photoprotective role in chlorophytes. While functional in diatoms, its pigments are present in small amounts compared to other photopigments, and mainly serve as precursors to the Ddx cycle pigments and Fx [Coesel et al. 2008, Lohr and Wilhelm 1999].

There are two major hypotheses regarding the biosynthesis of Ddx, Dtx, and Fx from Vx **(Fig. 2.1)**. Ddx and Fx may have a common precursor, or Fx may be synthesized from Ddx. Pigment flux-based evidence supports the latter hypothesis [Goericke and Welschmeyer 1992, Lohr and Wilhelm 1999], but it is yet to be directly tested by genetic manipulation, and enzymes involved in the biosynthetic steps have not been identified. Furthermore, Phaeophyceae and chrysophytes contain Fx but not Ddx cycle pigments, and therefore Ddx is not an obligate precursor of Fx [Lohr and Wilhelm 1999]. Neoxanthin has been hypothesized as an intermediate between Vx and Ddx/Fx based on its structure. Although typically not observed in diatom pigment extracts, it was detected in an enriched fraction from the diatom *Phaeodactylum tricornutum* by Dambeck et al. [2012].

The sequencing of diatom genomes and development of genetic manipulation tools have facilitated the study of diatom carotenoid biosynthesis. Putative carotenoid biosynthesis gene

candidates have been identified in the first two available diatom genomes, *Thalassiosira pseudonana* [Armbrust et al. 2004] and *P. tricornutum* [Bowler et al. 2008], by Coesel et al. [2008] and Dambeck et al. [2012], based on sequence identity of their protein products to those of known carotenoid biosynthesis genes. Functional complementation confirming enzyme functions for the pre-β-car part of the pathway was performed by Dambeck et al. [2012]. Lavaud et al. [2012] knocked down the *P. tricornutum* gene that encodes Vx de-epoxidase (VDE), which performs de-epoxidation in both xanthophyll cycles, resulting in a Dtx de-epoxidation deficiency (Vx de-epoxidation was not examined). Phytoene synthase (PSY), which catalyzes the first committed step of carotenoid biosynthesis, has been knocked down [Kaur and Spillane 2014] and overexpressed [Kadono et al. 2015] in *P. tricornutum*, resulting in reduced and increased cellular carotenoid content, respectively. Nevertheless, much remains to be understood about the carotenoid biosynthesis pathway in diatoms. In this study, we confirm and expand the list of putative carotenoid biosynthesis gene candidates using sequence identity-based reciprocal probing of the *T. pseudonana* and *P. tricornutum* genomes. We then focus on *T. pseudonana* and use available transcriptomic and physiological data to narrow down the carotenoid biosynthesis gene candidate list and form hypotheses about the functions of their protein products. Furthermore, we directly test the function of several enzymes hypothesized to be involved in diatom carotenoid biosynthesis through genetic manipulation.

2.3 RESULTS

2.3.1 Sequence Identity-Based Candidate Carotenoid Biosynthesis Gene Identification

The genomes of *T. pseudonana* and *P. tricornutum* were searched for genes that encode proteins with sequence similarity to those known to participate in carotenoid biosynthesis in both

organisms, in order to confirm what has been previously reported [Coesel et al. 2008, Dambeck et al. 2012] and identify new candidates. Some Basic Local Alignment Search Tool (BLAST) results indicated previously unreported gene models that overlap with previously published ones. The latter likely represent outdated models that have since been replaced. Large chromosomal pieces, sometimes encompassing genes identified as part of the carotenoid biosynthesis pathway or those found using the latter as BLAST queries, also frequently showed up in BLAST results. They may contain multiple partial sequence matches, possibly including those of domains that are present in, but not necessarily limited to, enzymes of the carotenoid biosynthesis pathway. Some open reading frames without available gene models were also found. The findings are detailed in **Appendix 2.A**, and the most current model IDs of known carotenoid biosynthesis genes and corresponding BLAST hits for which gene models are available are summarized in **Table 2.1**. The unnamed enzymes are not listed in a specific order and no relationship between those listed on the same line is implied. The diatom carotenoid biosynthesis pathway is depicted in **Fig. 2.1.**

2.3.2 Sequence Identity-Based Phylogenetic Analyses

Carotene cis/trans isomerase (CRTISO), LUT1-like (LTL), and zeaxanthin epoxidase (ZEP) candidate searches yielded numerous new candidates. Because many of the only publicly available Department of Energy Joint Genome Institute (DOE JGI) gene models for the LTL group are clearly incomplete, sequence identity analysis was not performed at this stage. However, it should be mentioned that the only BLAST hits for Phatr2_34027 in the *T. pseudonana* genome were Thaps3 9541 (LTL1) and Thaps3 2703365 (LTL2). Most enzymes in this group were functionally annotated as cytochrome p450, with some exceptions. Phatr2_47234, Phatr2_37006, Thaps3_4027, and Thaps3_4026 had no functional annotation available on the DOE JGI website. Phatr2_43563 was

functionally annotated as an iron-binding oxidoreductase, Phatr2_50619 chitinase, Phatr2_43537 chitinase with a diacylglycerol binding domain, Phatr2_43469 diacylglycerol acyltransferase, Phatr2_32833 N-acetyltransferase, and Thaps3_14875 glycosyltransferase (UDP-N-acetylglucosamine dolichyl-phosphate N-acetylglucosamine-phosphotransferase).

Sequence identity analysis for the CRTISO group **(Fig. 2.2A, Appendix 2.B1)** revealed that Thaps3_5221, Thaps3_21847, Thaps3_10233, Thaps3_21900, and Thaps3_7094 are more similar to the *P. tricornutum* enzymes identified by Dambeck et al. [2012] as CRTISO candidates, while Thaps3_25361, Thaps3_11636, Thaps3_5859, and Thaps_10254 grouped closer to each other and were more dissimilar from the former group. However, both groups had members that either had Gene Ontology (GO) terms including: carotenoid biosynthetic process, oxidoreductase activity, and flavin adenine dinucleotide (FAD) binding, or did not have GO terms available. All of the proteins had the EuKaryotic Orthologous Groups Identity (KOG ID) "phytoene desaturase."

ZEP candidate sequence identity analysis **(Fig. 2.2B, Appendix 2.B2)** showed that previously identified ZEPs [Coesel et al. 2008] form a group without additional members. Within the group, the ZEP1s and ZEP2s clustered with their counterparts, and ZEP3 was separate. No terms, predictions, or annotations were available for Phatr2_48545 (ZEP1). All of the other proteins that had a KOG ID were designated as kynurenine 3-monooxygenase and related flavoprotein monooxygenases, and the two that did not were indicated to have a FAD-binding domain (Phatr2_56492, ZEP3) or to be a FAD-dependent oxidoreductase (Phatr2_56488, ZEP2) by Pfam. All of the proteins besides Phatr2_48545 (ZEP1) had GO terms that included monooxygenase activity and cellular aromatic compound metabolic process. Most also included oxidoreductase activity, with the exception of Thaps3_22671 and Phatr2_43425. GO terms for Phatr2_45936 also included small nuclear ribonucleoprotein complex, and for Thaps3_261390 (ZEP2), xanthophyll biosynthetic process.

Sequence identity analysis was also performed on the VDE/VDL/VDR group **(Fig. 2.2C, Appendix 2.B3)**. Four clusters resulted, with the known VDEs, VDL1s, and VDRs clustering with their counterparts, and the *P. tricornutum* VDL2 in a cluster with the two newly identified proteins, Phatr2_bd_1281 and Thaps3_11707. Based on the sequence similarity, the latter is hereafter designated as the *T. pseudonana* VDL2.

2.3.3 Candidate Gene Expression Patterns

For all genes encoding the proteins identified in 2.3.1 **(Table 2.1)**, expression patterns were obtained from the microarray dataset described in Smith et al. [2016]. Briefly, Smith et al. performed a 24-hour silicon starvation time course experiment, with cells placed in silica-free media at 0 h, sampling for a variety of physiological variables as well as the transcriptome throughout. At 4 h, the majority of cells underwent chloroplast division. There was also evidence of sustained light-induced stress throughout the duration of the experiment. Genes with similar functions appeared to be highly co-regulated, as evidenced by their sorting into clusters with similar expression patterns.

Most genes known to participate in *T. pseudonana* carotenoid biosynthesis followed a distinct expression pattern that spiked at 4 h, during chloroplast division, then decreased **(Fig. 2.3A)**. Some increased expression by 4 h and remained elevated throughout, and ZEP2 was unique in having lower transcript levels throughout the experiment compared to 0 h **(Fig. 2.3B)**. One half (15 out of 30) of the candidate genes identified in 2.3.1 **(Table 2.1)** based on sequence identity had expression patterns similar to those known to be in the carotenoid biosynthesis pathway **(Fig. 2.3C, D)**, the other half did not **(Fig. 2.3E, F)**.

Thaps3_17707, the novel VDL2 (2.3.2), and Thaps3_10233, found in the sequence identity-based search for CRTISO candidates **(Table 2.1)**, were upregulated to a substantially higher extent than the rest of the carotenoid biosynthesis genes **(Fig. 2.3A-D)** and were previously found by Smith et al. [2016] to be in a co-regulated gene cluster with photoantenna protein-encoding genes. We examined the expression patterns of those photoantenna protein genes as well as carotenoid biosynthesis genes using an additional set of transcriptomic data [Abbriano 2017]. This RNA-seq data was obtained during a 9-hour time course experiment after re-addition of silica to silicon-starved *T. pseudonana* cultures, which typically results in a synchronous progression through the cell cycle and thus synchronous chloroplast division. VDL2 and Thaps3_10233 had a distinct co-expression pattern in the RNA-seq silicon re-addition as well and were upregulated later and to a higher extent than the rest of the carotenoid biosynthesis genes **(Fig. 2.4A)**. All but one photoantenna protein-encoding gene that clustered with VDL2 and Thaps3_10233 in the microarray dataset [Smith et al. 2016] also clustered with them in RNA-seq dataset [Abbriano 2017] **(Fig. 2.4B)**.

2.3.4 Full-length Gene Models and Predicted Protein Targeting

Because gene models on the DOE JGI website are sometimes incomplete/incorrect, full-length gene models **(Appendix 2.C)** were constructed by using available RNA-seq data [Abbriano 2017, Smith et al. 2016] for all the genes known to be part of the carotenoid biosynthesis pathway, those that had similar microarray gene expression patterns [Smith et al. 2016], and Thaps3_263437 ("β-carotene hydroxylase, (BCH)"), which did not **(Fig. 2.3)**. Full-length gene models were translated **(Appendix 2.C)**, and the likelihood of chloroplast targeting for the resultant peptides was assessed **(Appendix 2.C)**, summarized in **Table 2.2**.

2.3.5 Additional Sequence-Based Analyses

2.3.5.1 Phytoene Desaturase (PDS) and Thaps3_bd_1474

Thaps3_bd_1474, found in the PDS candidate search **(Table 2.1, Appendix 2.A)**, is part of the unmapped sequence assembly of the *T. pseudonana* genome and is truncated on the 5' end due to the way those sequences were assembled. The unmapped sequence assembly is also referred to as the "bottom drawer," and "_bd_" is added to the identification numbers of genes that are part of it. RNA-seq data is not available for the "bottom drawer" sequences and could not be used to construct a full-length gene model. The JGI predictions were used instead **(Appendix 2.D)**. The JGI-predicted peptides for PDS1 and Thaps3_bd_1474 are identical where they align. The latter is approximately 200 amino acids shorter on the N-terminal side and approximately 120 amino acids longer on the C-terminal side. The PDS1 gene contains no introns, and one intron is predicted by JGI for Thaps3_bd_1474. If the predicted intron is disregarded and the sequence is translated as one open reading frame **(Appendix 2.D)**, the resulting peptide is identical to PDS1, except for the N-terminal truncation **(Appendix 2.D)**. When genomic DNA sequences that include the JGI models for PDS1 and Thaps3_1474 and 10 kb downstream are examined, they appear identical where the gene models align and for approximately 1.8 kb downstream, with occasional single base pair differences that may be attributed to naturally occurring polymorphisms, and then abruptly become substantially different thereafter **(Appendix 2.D)**. The identical region downstream PDS1 and Thaps3_bd_1474 includes two more genes, one predicted to encode a ribosomal protein L33 (Thaps3_41211 and Thaps3_bd_1472), and the other predicted to encode a DNA methylase (Thaps3_6523 and Thaps3_bd_1109). By contrast, PDS1 and PDS2 have 62.9% sequence identity **(Appendix 2.D)**.

2.3.5.2 Thaps3_263437 ("β-carotene Hydroxylase, BCH")

Out of the two major known BCH types [Martin et al. 2008], Thaps3_263437 was found to be more similar to the non-heme di-iron enzymes than to the cytochrome P450 monooxygenases **(Appendix 2.D)**. Approximately one half of the *T. pseudonana* protein, on the N-terminal side, aligns with known BCH sequences from other organisms. The Thaps3_263437 protein sequence was compared to 10 non-heme di-iron enzymes from diverse organisms, including two plants, one green alga, three cyanobacteria, three non-photosynthetic eubacteria, and one archaeal species, which represent four major groups of the non-heme di-iron BCHs [Tian and DellaPenna 2004]. The least similarity was found with cyanobacteria, ranging from 17-22% sequence identity. The sequence identity range for the other examined organisms is between 26-29% **(Appendix 2.D)**. No known motifs or conserved domains based on sequence or predicted structure were found in the C-terminal half of Thaps3_263437. Portions of it did exhibit limited sequence identity to unknown proteins from diverse organisms **(Table 2.3)**. Neither the N-terminal ("BCH") half nor the unknown C-terminal half of Thaps3_263437 were found in the other currently available diatom genomes (*Phaeodactylum tricornutum, Cyclotella cryptica*, *Thalassiosira oceanica*, *Fragilariopsis cylindrus*, *Pseudo-nitzschia multiseries*).

2.3.5.3 Functional Annotation Predictions for the T. pseudonana Carotenoid Biosynthesis Enzymes

Sequence-based functional annotation predictions were compiled for the enzymes of the *T. pseudonana* carotenoid biosynthesis pathway **(Table 2.4)**. Additionally, predicted protein structures for VDL2, Thaps3_10233, and the C-terminal half of Thaps3_263437 were analyzed for functional predictions. The analysis yielded no additional information.

2.3.6 Identification of Carotenoid Biosynthesis Genes in Additional Diatom Species

Orthologs of the carotenoid biosynthesis genes identified in *T. pseudonana* and *P. tricornutum* (2.3.1-2.3.4) were found in the four other currently available diatom genomes **(Table 2.5)**. Where gene models were not available, genomic coordinates were given. Three of the *T. oceanica* genes appeared to be split into two adjacent gene models. Those were listed together, separated by a plus sign (+). The number of copies of each gene varied between species, with the most for several genes found in *C. cryptica* **(Table 2.5)**.

2.3.7 Genetic Manipulation and Resultant Pigment Phenotypes

The following transgenic *T. pseudonana* lines were created to investigate the roles of several enzymes hypothesized to participate in carotenoid biosynthesis: overexpression (OE) of the novel VDL2, knockdown (KD) of VDL1, KD of VDL2, and a simultaneous KD of LUT1-like (LTL) 1 and LTL2, which share 48% sequence identity **(Appendix 2.D)**. Four independent clones per line were then allowed to adapt to cultivation conditions, and their photosynthetic pigment content was assessed using high-performance liquid chromatography (HPLC).

For VDL2 OE, total cellular photosynthetic pigment content (Tot) varied between the four clones and two WT cultures at both low (30 µmol photons m^{-2} sec^{-1}) and high light (300 µmol photons m^{-2} sec^{-1}) without an observable trend **(Fig. 2.5A)**. Nevertheless, certain pigment ratios exhibited a consistent difference between WT and VDL2 OE. Fx/Tot was increased (low light p-value = 0.02, high light p-value = 0.0008) and (Ddx+Dtx)/Tot was reduced (low light p-value = 0.0006, high light p-value = 0.0008) in VDL2 OE, but (Ddx+Dtx+Fx)/Tot was not significantly different from WT (low light p-value = 0.9, high light p-value = 0.8) **(Fig. 2.5B)**. No differences were observed in β-car/Tot, Chl a/Tot, and Chl c/Tot **(Fig. 2.5C, D)**.

For LTL KD lines, initial PCR-based screening for the presence of the full knockdown construct (promoter to terminator) revealed a selection against it, as most of the screened clones had integrated the portion allowing them to grow on selection but lost all or parts of the promoter and/or the antisense portion. Approximately 3 weeks after plating, significant pigmentation reduction became apparent in some of the clones **(Fig. 2.6A)**. Four LTL KD clones were selected based on lighter pigmentation, and genomic integration of the entire KD construct was subsequently confirmed by PCR. There was a reduction in Tot in LTL KDs compared to WT, with more pronounced differences in high light than in low light **(Fig. 2.6B)**. In low light, LTL KD lines had 75-93% Tot of average WT (p-value = 0.2), and 60-76% of average WT Tot in high light (p-value = 0.009). No statistically significant differences in the ratios of individual photosynthetic pigments to Tot were found **(Fig. 2.6B-D)**.

Four clones each of VDL1 KD and VDL2 KD were selected based on the ability to grow on selection and PCR-assessed genomic integration of full-length knockdown constructs. Photosynthetic pigment content was assessed in high light only and Tot was found to be reduced in both lines, with the exception of the clone VLD2 KD3 **(Figs. 2.7A, 2.8A)**. VDL1 KD lines contained 48-91% of the average WT Tot (p-value = 0.01), and VDL2 KD lines were at 59-87% of average WT Tot (p-value = 0.004), excluding KD3. KD3 did not significantly differ from WT in Tot. No substantial differences in the ratios of individual pigments to Tot were observed between WT and the VDL1 or VDL2 KD lines **(Figs. 2.7, 2.8)**.

2.4 DISCUSSION

2.4.1 Carotenoid Biosynthesis Candidate Gene/Enzyme Identification

Building on earlier analyses [Coesel et al. 2008, Dambeck et al. 2012], we sought to identify novel gene candidates based on sequence identity of their protein products to those of previously identified carotenoid biosynthesis genes. The genomes of *T. pseudonana* [Armbrust et al. 2004] and *P. tricornutum* [Bowler et al. 2008] were chosen for reciprocal BLAST searches as the first two sequenced, and best annotated, available diatom genomes. A notable finding was that of an additional, previously unreported VDL protein in *T. pseudonana* (Thaps3_11707), designated VDL2 in this study. It should be noted that the *T. pseudonana* VDR (Thaps3_270211) is mis-annotated as VDL2 on the DOE JGI website.

The list of candidates for the *T. pseudonana* carotenoid biosynthesis pathway obtained by BLAST was narrowed down to those whose gene expression patterns were similar to those of genes known to participate in carotenoid biosynthesis, using available transcriptomic (microarray) data [Smith et al. 2016] **(Fig. 2.3)**. The data was obtained from a 24-hour silicon starvation time course during which there was a major chloroplast division event at 4 h, and light-induced stress was sustained throughout. The latter is at least partially attributable to the culture being resuspended in silica-free media at half the density that it had grown to in regular media prior to being harvested, resulting in reduced shading [Smith et al. 2016]. Both chloroplast division and exposure to higher light are expected to result in carotenoid biosynthesis [Chapter 1, Lepetit et al. 2010], and Smith et al. [2016] did measure an overall steady increase in photopigment content throughout the time course, as well as a preferential accumulation of the photoprotective de-epoxide form (Dtx) of the interconvertible Ddx/Dtx cycle pigments. Thus, carotenoid biosynthesis activities associated with both chloroplast division and light-induced stress occurred during the experiment. Smith et al.

[2016] report substantial co-regulation of genes with related functions, as evidenced by numerous gene clusters with highly similar expression patterns. Nevertheless, it is possible that some gene candidates whose expression patterns differed from those of known carotenoid biosynthesis genes and were therefore not chosen for further study **(Fig. 2.3),** may be relevant to carotenoid biosynthesis as well, but regulated differently. Smith et al. [2016] also collected RNA-seq data alongside the microarray data, with a notable difference of the culture used for the former being resuspended in silica-free media at the same density that it had achieved in regular media prior to harvesting. Thus, unlike the culture used for the microarray data, the culture used for RNA-seq would not have been exposed to more light due to reduced shading at the beginning of the silicon starvation experiment. Even though expression patterns for many genes replicated well between the microarray and RNA-seq silicon starvation datasets, those of genes known to be involved in the light-stress response [Smith et al. 2016], as well as those of genes known to participate in carotenoid biosynthesis **(Figs. 2.3, S2.1)** did not, suggesting that the responses observed in the microarray dataset could not be attributed solely to silicon starvation but were affected by increased light exposure of the microarray culture as well. In summary, even though silicon starvation is not expected to directly affect carotenoid biosynthesis, the microarray data obtained by Smith et al. was relevant to our study. Synchronized chloroplast division and light-induced stress occurred during their experiment, requiring distinct carotenoid biosynthesis pathway responses. The silicon starvation RNA-seq dataset [Smith et al. 2016] serves as a negative control, as it was obtained from a culture treated the same way as the one used for the microarray data, except being exposed to the conditions that induced the light-induced stress response. The fact that most genes had similar expression patterns in both datasets, but light-stress response [Smith et al. 2016] and carotenoid biosynthesis genes did not **(Figs. 2.3, S2.1)**, highlights the relevance of the microarray dataset to our study. No physiological data was collected from the silicon starvation

RNA-seq culture, and thus the status of light stress and pigment abundance cannot be directly referenced for that dataset [Smith et al. 2016].

Candidates were further narrowed down to those predicted to be targeted to the chloroplast by using RNA-seq-derived [Abbriano 2017, Smith et al. 2016] full-length gene models **(Appendix 2.C, Table 2.2)**. Due to diatoms having evolved via secondary endosymbiosis, their chloroplasts are surrounded by the endoplasmic reticulum (ER) [Smith et al. 2012]. Thus, diatom proteins targeted to the chloroplast must also be targeted to the ER. PSY2 and LTL2 had clear ER targeting, but their predictive scores for chloroplast targeting were just below the cutoff value **(Appendix 2.C, Table 2.2)**. It is likely that they are indeed targeted to the chloroplast, as chloroplast targeting predictions are not always completely accurate [Emanuelsson et al. 1999].

The majority of candidates identified by BLAST, including the numerous LTL, CRTISO, and ZEP candidates, did not meet the criteria of both having microarray gene expression patterns similar to genes known to participate in carotenoid biosynthesis and predicted chloroplast targeting. Barring potentially having excluded novel carotenoid biosynthesis genes that had different gene expression patterns, this indicates that most of the candidates were found due to containing domains that are utilized by, but not limited to, carotenoid biosynthesis enzymes. Indeed, each of the CRTISO, LTL, and ZEP candidate groups had shared functional annotation predictions.

2.4.2 Stepwise Analysis of the T. pseudonana Carotenoid Biosynthesis Pathway

The condensation of two geranylgeranyl pyrophosphate molecules to synthesize phytoene by PSY is the first committed step of carotenoid biosynthesis and is considered rate-limiting [Bertrand 2010, Eilers et al. 2016]. It is followed by several desaturation steps catalyzed by PDS [Melendez-Martinez et al. 2015]. There are two previously reported copies of each in *T. pseudonana* [Coesel et al. 2008], and a possible additional *T. pseudonana* PDS identified in this study. Barring mis-assembly, the novel PDS candidate Thaps3_bd_1474 appears to have arisen as a result of a duplication event of the PDS1 locus, encompassing it, as well as two downstream genes (2.3.5.1). Because of the way "bottom drawer" sequences were processed, the N-terminus, as well as the genomic sequence upstream of Thaps3_bd_1474 are not available, and the size of the duplicated region is not known.

Both PSY and PDS genes exhibited two distinct gene expression patterns in the microarray data [Smith et al. 2016]. PSY1, PDS1, and Thaps3_bd_1474 had a spike in expression at 4 h, during the major chloroplast division event **(Fig. 2.3A, C)**. By contrast, PSY2 and PDS2 had their transcript levels increase by 4 h and stay elevated throughout the time course **(Fig. 2.3B)**. In vascular plants that possess more than one copy of PSY, the different genes are differentially regulated in response to developmental and environmental signals. Differential regulation based on different cellular needs may be hypothesized for microalgae that possess more than one PSY gene, but this has yet to be experimentally investigated [Melendez-Martinez et al. 2015, Tran et al. 2009]. We hypothesize that in *T. pseudonana*, PSY1 may serve to initiate carotenoid biosynthesis during chloroplast division and PSY2 may serve as an independent way to initiate carotenoid biosynthesis in response to increased irradiance, which leads to an increase in Ddx cycle pigments [Lavaud et al. 2004]. Because PDS is responsible for the next immediate steps in carotenoid biosynthesis and has also been previously suggested to be rate-limiting [Chamovitz et al. 1993], different PDS copies might be

differentially co-regulated with different copies of PSY. As detailed in Chapter 1, the cellular accumulation of all diatom photosynthetic pigments is influenced by cell cycle progression and therefore chloroplast division, with the exception of the Ddx cycle pigments, which appear to primarily respond to light intensity. The differential accumulation of Ddx+Dtx and Fx, the other end product of carotenoid biosynthesis, as well as β-car, a precursor for all three aforementioned pigments, suggests a differential regulation mechanism, such as independent biosynthesis induction by different copies of PSY/PDS, as proposed above. PSY1, PDS1, and Thaps3_bd_1474 are upregulated during chloroplast division along with the majority of carotenoid biosynthesis genes **(Fig. 2.3)** as well as chlorophyll biosynthesis genes and photoantenna protein-encoding genes [Smith et al. 2016], and thus likely respond to the cell's need to populate dividing chloroplasts with enzymes and pigments. PSY2 and PDS2 display a gene expression pattern similar to that of PsbA **(Fig. 2.3B)**, a gene that encodes the photosystem II D1 protein, which is known to turn over in response to light-induced stress [Domingues et al. 2012, Wu et al. 2011]. Thus, PSY2 and PDS2 transcript levels may have been elevated throughout the time course to activate carotenoid biosynthesis in order to support the accumulation of Ddx+Dtx in response to the light-induced stress that was observed throughout the experiment, independent of the major chloroplast division event at 4 h. The transcript responses for PSY1/PDS1 and PSY2/PDS2 showed similarity in the RNA-seq data set by Smith et al. [2016] as well **(Fig. S2.1A)**.

Following PDS, more desaturation is carried out by ζ-carotene desaturase (ZDS), resulting in pro-lycopene, a lycopene stereoisomer. There is only one known copy of ZDS in *T. pseudonana*, and we present no new hypotheses or observations for this step. Pro-lycopene is isomerized to lycopene by CRTISO, and the identity of the enzyme(s) responsible for this step is not yet confirmed. Many potential CRTISO copies have been previously found in diatom genomes, including *T. pseudonana* [Bertrand 2010]. Our analysis has narrowed the list down to two candidates,

Thaps3_21900 and Thaps3_10233 **(Fig. 2.2A, Table 2.2)**. It is possible, however, that other candidates identified by BLAST **(Table 2.1, Appendix 2.A)** might participate in this step but were excluded based on having a gene expression pattern that differed from genes known to participate in carotenoid biosynthesis **(Fig. 2.3)**, or due to a false negative prediction for chloroplast targeting **(Table 2.2)**.

Lycopene is converted into β-car by lycopene cyclase B (LCYB), for which no new candidates or hypotheses were generated by this study. The subsequent conversion of β-car via hydroxylation to Zx is intriguing. That reaction is typically catalyzed by one of the two known types of BCH [Martin et al. 2008], neither of which are found in any of the currently available diatom genomes, except a partial 238aa sequence in *T. pseudonana* (2.3.5.2). The latter, however, is not predicted to be targeted to the chloroplast, and has a 342aa C-terminal addition which has sequence identity to unidentified peptides from several diverse organisms **(Table 2.3)**. Its gene expression pattern in the microarray dataset [Smith et al. 2016] is markedly different from any known carotenoid biosynthesis genes **(Fig. 2.3E)**. Like the BCH sequence, the C-terminal peptide is not found in any other currently available diatom genomes (*C. cryptica*, *T. oceanica*, *F. cylindrus*, *P. multiseries*, *P. tricornutum*), which makes the *T. pseudonana* BCH an evolutionary curiosity. It is unlikely to participate in carotenoid biosynthesis and has possibly evolved to perform a different function by losing chloroplast targeting and gaining the C-terminal sequence.

Diatom genomes contain genes encoding LUT1-like (LTL) enzymes, which are similar to an *Arabidopsis thaliana* enzyme (LUT1) that converts α-car to lutein (respective isomers of β-car and Zx) and has been demonstrated to have weak β-car hydroxylation activity. Since diatoms do not make α-car, it has been hypothesized that diatom LTLs function in place of BCH, but this has not been previously experimentally assessed [Bertrand 2010]. To do so, we simultaneously knocked down both *T. pseudonana* LTL copies, which resulted in a reduction of all photosynthetic pigments

(Fig. 2.6). Although this does not conclusively place the LTLs in a specific carotenoid biosynthesis step, it does implicate them as being involved. Both LTLs are cytochrome P450 monooxygenases **(Table 2.4)**, and the hydroxylation of β-car to produce Zx is the first step in the carotenoid biosynthesis pathway in which oxygen atoms are added **(Fig. 2.1)**. Thus, it is unlikely that the LTLs catalyze anything upstream of β-car hydroxylation.

Zx, whether synthesized from β-car by the LTLs or another enzyme that is yet to be discovered, is part of the Vx xanthophyll cycle. The de-epoxidation reactions of both the Vx and the Ddx xanthophyll cycles are catalyzed by ZEPs [Bertrand 2010, Goss et al. 2006, Wilhelm et al. 2006]. Chlorophytes lack the Ddx cycle and typically possess one ZEP copy, while multiple ZEPs are found in all diatom genomes available to date **(Table 2.5)**. It has been demonstrated that Ddx de-epoxidation is strongly inhibited by the transthylakoid proton gradient (ΔpH) that forms in high light. The ZEPs that operate in the chlorophyte Vx cycle, by contrast, are not inhibited by high light [Goss et al. 2006, Wilhelm et al. 2006]. Experimental evidence suggests that Zx epoxidation proceeds under high light in diatoms as well, since they perform *de novo* biosynthesis of the Ddx cycle pigments upon an increase in illumination [Lavaud et al. 2004, Lepetit et al. 2010], for which Vx is a precursor [Lohr and Wilhelm 1999]. Thus, different copies of diatom ZEPs may distinctly participate in the two xanthophyll cycles. We found that the two *T. pseudonana* ZEP copies had distinct gene expression patterns in the microarray data [Smith et al. 2006] **(Fig. 2.3A, B)**. Throughout the experiment, the abundance of light-harvesting and photoprotective pigments increased, and the Ddx pool pigments continuously preferentially accumulated as Dtx, the de-epoxide [Smith et al. 2006]. Thus, we hypothesize that ZEP2, downregulated throughout the experiment, participates in the Ddx cycle, whereas ZEP1, upregulated especially during the major chloroplast division event at 4 h, participates in the Vx cycle.

De-epoxidation of the Vx cycle pigments in chlorophytes is catalyzed by VDE. Only one VDE copy is found in the available diatom genomes **(Table 2.5)**, and it has been demonstrated that it participates in both diatom xanthophyll cycles [Jakob et al. 2000, Lavaud et al. 2012]. In addition to the VDE, diatom genomes encode VDE-like (VDL) and VDE-related (VDR) proteins **(Fig. 2.2C)**, the function of which has not been previously determined. It has been stipulated that they may have xanthophyll cycle activity in addition to the VDE, and perhaps differ in localization from the latter [Bertrand 2010, Lavaud et al. 2012]. Because VDR appears to be present in all chlorophytes [Coesel et al. 2008], it is unlikely to catalyze any reactions that are unique to diatoms. In the present study, we identified a previously unreported VDL2 in *T. pseudonana* (Thaps3_11707) based on sequence identity to the VDL2 in *P. tricornutum*. Overexpressing it resulted in increased cellular Fx abundance, together with a stoichiometric decrease in the cellular abundance of Ddx+Dtx. The ratio of Ddx+Dtx+Fx to total cellular pigments remained unchanged **(Fig. 2.5)**. Since the abundance of Chl a did not increase along with Fx, we suggest that the excess Fx in the VDL2 OE lines was either localized in the lipid shield around the photoantennae, along with non-protein bound Ddx cycle pigments, or that it may have replaced some of the photoantenna-bound Ddx [Lepetit et al. 2010]. Eilers et al. [2015] also found that increased Fx abundance in their *P. tricornutum* PSY OE lines was not accompanied by a concomitant increase in Chl a content. While our results do not clarify whether Ddx and Fx share a common precursor or Fx is derived from Ddx, they do implicate VDL2 in Fx biosynthesis. This is the first report of an enzyme being involved in Fx biosynthesis, and the first experimental evidence for a VDL function. Our findings do not support the hypothesis that VDLs participate in xanthophyll cycling. Multiple chemical reactions must take place during Fx biosynthesis from either Ddx or the other hypothetical precursor, neoxanthin. It is possible that VDL2 is not the sole enzyme responsible for Fx biosynthesis. However, based on our findings, if other enzymes are involved in this step, they are not rate-limiting. VDL2 analysis based on sequence

as well as predicted structure revealed only a VDE-like lipocalin domain, a fold that is shared by VDEs, VDLs, and ZEPs, and allows them to bind to their pigment substrates [Coesel et al. 2008]. The chemistry necessary to synthesize Fx includes additions of one acetyl and one keto group and oxidoreductase activity. It is possible that VDL2 is able to catalyze all of the required reactions. It is capable of oxidoreductase activity **(Table 2.4)**, which could catalyze everything except for the acetyl group addition. It should be noted that VDL2 is designated as part of the N-acetylglucosaminyltransferase KOG group on the DOE JGI website **(Table 2.4)**, but it is unclear why the software made this prediction. No corresponding domains have been identified either by VDL2 sequence and predicted structure analyses nor by aligning VDL2 to other proteins in that group, which yielded only very sparse similarity not limited to any specific part of the sequence (data not shown). If the software designation is accurate, however, it may indicate an ability to transfer acetyl groups. Almost half of the VDL2 protein on the N-terminal side shows no similarity to other known proteins based on either sequence or predicted structure and may perform previously undescribed chemistry.

Like VDL2, Thaps3_10233, an oxidoreductase **(Table 2.4)** initially identified by BLAST during the CRTISO candidate search, exhibited gene expression patterns in both the microarray dataset [Smith et al. 2016] and RNA-seq dataset [Abbriano 2017] that were distinct from the other carotenoid biosynthesis genes and similar to photoantenna protein genes **(Figs. 2.3, 2.4)**. We suggest that Thaps3_10233 may be involved in Fx biosynthesis along with VDL2. The co-expression of the genes encoding proteins involved in Fx biosynthesis with photoantenna protein genes may represent another layer of carotenoid biosynthesis regulation, in addition to the hypothesized separate means of inducing the pathway in response to irradiance increase and chloroplast division. Synthesizing Fx only when photoantenna proteins are also being made may be a way to ensure that pigment precursors are funneled into Fx biosynthesis only when it is needed to populate newly

synthesized photoantenna proteins, and otherwise are used to make Ddx cycle pigments. While some Ddx cycle pigments are bound to photoantenna proteins, the majority of those synthesized in response to an increase in illumination are dissolved in the lipid shield around photoantennae [Lepetit et al. 2010]. Therefore, their accumulation would not need to be coordinated with photoantenna protein production. The hypothesized mechanisms of differential carotenoid biosynthesis regulation are summarized in **Fig. 2.9**.

2.4.3 Total Photopigment Reduction in Knockdown Lines

An overall reduction in cellular photopigment content was observed in LTL, VDL1, and VDL2 KD lines, without substantial differences in the ratios of individual pigments **(Figs. 2.6-2.8)**. Thus, β-car, presumably upstream of the biosynthetic steps targeted by the KDs, as well as Chl a and Chl c, which are part of a separate biosynthetic pathway, were reduced proportionately with Fx and Ddx cycle pigments in the KD lines. Carotenoids are known to play a crucial role in the assembly and stabilization of light-harvesting complexes in photosynthetic bacteria, algae, and green plants [Moskalenko and Karapetyan 1996, Santabarbara et al. 2013]. Unlike chlorophytes that adjust the size of photoantenna complexes associated with photosynthetic reaction centers in response to changes in light intensity, diatoms co-regulate the number of reaction centers and photoantenna units. Thus, the ratio of Chl a to Fx and β-car does not vary substantially, while Chl c abundance appears to be more variable [Brunetet al. 2014, Lepetit et al. 2012]. In the KD lines, Chl a, Chl c, and β-car may have failed to accumulate due to being prevented from stable binding to light-harvesting complex proteins because of light-harvesting complex destabilization caused by a Fx deficit, especially since the pigment ratios appear to be generally maintained in diatoms. Lohr and Wilhelm [1999] monitored photopigment accumulation in *P. tricornutum* upon reducing illumination

intensity, with and without the addition of the *de novo* carotenoid biosynthesis inhibitor norflurazon, which acts upon PDS. In the norflurazon-treated sample, the abundance of Fx increased as accumulated precursor pigments were depleted. In the untreated samples, following the initial conversion of precursors to Fx, Fx abundance continued to increase, presumably via *de novo* biosynthesis. Despite the differences in final Fx content between the treated and untreated samples, final Chl a/Fx and β-car/Fx ratios did not differ substantially, demonstrating that Chl a and β-car accumulation was proportional to that of Fx. Chl c abundance was not reported [Lohr and Wilhelm 1999]. This observation provides further support for the notion that reducing the abundance of Fx will also reduce that of the other photopigments, as was observed in our KD lines **(Figs. 2.6-2.8)**.

2.4.4 Broader Implications of the Findings

Our results have important implications for studying diatom carotenoid biosynthesis. If knocking down most steps in the pathway will result in the same phenotype of overall pigment reduction, it will not allow the elucidation of enzyme function, as was the case with all knockdowns performed in our study. However, if it is possible to perform chemical rescue experiments by feeding various pigments in the pathway to knockdown lines, that may be a viable approach to further pathway elucidation. There are some exceptions, however, for which the knockdown approach alone may be helpful. VDE, for example, does not participate in forward *de novo* carotenoid biosynthesis. It has been knocked down in *P. tricornutum* without reducing pigment content, resulting only in impaired Ddx de-epoxidation (and likely that of Vx as well, which was not measured) [Lavaud et al. 2012]. Knocking down the two *T. pseudonana* ZEPs may help confirm our hypotheses about them as well. If, as we expect, ZEP1 participates in the Vx cycle, knocking it down

should result in overall pigmentation reduction, as forward biosynthesis through that pathway is necessary for Vx biosynthesis, and therefore that of Fx and Ddx cycle pigments. However, if the hypothesis put forth by Lohr and Wilhelm [2001] that Vx is synthesized via β-cryptoxanthin from β-car is correct, the ZEP1 knockdown may not have as much of an effect on the downstream pigments, but may result in Zx accumulation under certain conditions [Lohr and Wilhelm 1999]. Knocking down ZEP2 on the other hand should not affect total cellular pigment content but rather result in a Dtx epoxidation defect upon a shift from high light to low light, if our hypothesis that it participates in the Ddx cycle is correct. Our hypothesis about there being two ways to independently induce carotenoid biosynthesis in response to either chloroplast division or an increase in illumination may be tested by knocking down PSY1 and PSY2 separately as well. If PSY1 indeed serves to initiate carotenoid biosynthesis during chloroplast division in order to populate newly divided chloroplasts with pigments, knocking it down should result in an overall pigment reduction. If PSY2 initiates carotenoid biosynthesis in order to make more Ddx cycle pigments in response to increased illumination, knocking it down should not affect cellular pigmentation at low light intensities, but should result in impaired Ddx cycle pigment pool accumulation upon being transferred to higher light. The aforementioned experiments were out of the scope of this study.

Overexpression, as a different approach to assessing enzyme function, may be limited by substrate availability and not usable for all the pathway steps. Heterologous complementation, such as performed by Dambeck et al. [2012] for the earlier carotenoid biosynthesis steps in *P. tricornutum*, may help in the study of the function of enzymes involved in the later steps as well.

Finally, our findings must be taken into consideration if altering diatom pigment content by genetically manipulating carotenoid biosynthesis is of interest. It does not appear possible to reduce the amount of the main accessory photopigment Fx without affecting the abundance of other photopigments, in order to accumulate β-car for commercial purposes or obtain a

photoantenna reduction phenotype that has been achieved in chlorophytes [e.g., Kirst et al. 2012], for example. On the other hand, it is possible to increase cellular Fx content, which may be of interest due to its various health-promoting activities [Peng et al. 2011].

The *T. pseudonana* genes identified and discussed in this study are also found in all other currently available diatom genomes **(Table 2.5)**. Thus, our findings are relevant to diatom carotenoid biosynthesis in general, and not limited to our model species. There may exist, however, interspecies differences in certain aspects of the pathway. For example, *P. tricornutum* has only one copy of PSY, unlike *T. pseudonana* and other diatoms with sequenced genomes available that have at least two **(Table 2.5)**. Therefore, the step that PSY catalyzes could not be used by *P. tricornutum* to differentially initiate carotenoid biosynthesis in response to chloroplast division and higher illumination. However, *P. tricornutum* does have two copies of PDS, which we hypothesize are differentially used by *T. pseudonana* along with the two PSY copies **(Fig. 2.9)**. Consequently, *P. tricornutum* might use PDS, but not PSY, to differentiate between the two possible carotenogenic needs of the cell. Another observation is that some carotenoid biosynthesis genes in *P. tricornutum* are adjacent and appear to be divergently transcribed [Coesel et al. 2008], whereas no such relationships exist between the carotenoid biosynthesis genes in *T. pseudonana*. Interestingly, two such pairs in *P. tricornutum* are VDE with ZEP3 and VDL2 with ZEP1. Since de-epoxidation and epoxidation of xanthophyll cycle pigments are in opposition of each other, it is possible that ZEP3 functions in a process other than xanthophyll cycling. *P. tricornutum* has three copies of ZEP **(Table 2.5)**, and it is possible that ZEP1 and ZEP2 differentially participate in the two xanthophyll cycles as we propose they do in *T. pseudonana* **(Figs. 2.1, 2.9)**, while the third copy has evolved to perform a different role. Xanthophyll epoxidation and Fx biosynthesis on the other hand are likely to co-occur, and thus the co-regulation of ZEP1 and VDL2 is not surprising. ZEP1 in *P. tricornutum* may participate in the Vx cycle as we suggest the *T. pseudonana* ZEP1 does, and upregulating ZEP1

together with VDL2 when Fx biosynthesis is necessary would be logical. Diatoms are incredibly diverse and adapted to a wide variety of environmental conditions over the course of evolution, and many strategies employed by different species may vary [Hildebrand et al. 2012], including those related to photosynthesis and photoprotection [Lavaud and Lepetit 2013]. However, increasing what is known about a process in one diatom species ultimately facilitates future efforts for studying it in others, and improves our overall understanding of these environmentally important and commercially promising organisms.

2.5 METHODS

2.5.1 Sequence Identity-Based Carotenoid Biosynthesis Gene Candidate Search

P. tricornutum and *T. pseudonana* genome sequence data was accessed on the Department of Energy Joint Genome Institute (DOE JGI) website at https://genome.jgi.doe.gov/Phatr2/Phatr2.home.html and https://genome.jgi.doe.gov/Thaps3/Thaps3.home.html, respectively. Database searches were performed using the tBLASTn function available on the website, comparing protein queries to translated nucleotide sequences. Redundant gene models, and those representing pieces of more complete models, were not listed. Sequence identity analyses were performed on protein sequences predicted by the DOE JGI website. Multiple sequence alignment, percent identity matrix generation, and phylogenetic tree construction were completed using Clustal Omega (https://www.ebi.ac.uk/ Tools/msa/clustalo/) [Sievers et al. 2011]. Trees were visualized using Interactive Tree of Life (iTOL) (https://itol.embl.de/) [Letunic and Bork 2016]. Gene Ontology (GO) terms, EuKaryotic Orthologous Groups Identity (KOG ID), and Protein Family (Pfam) information were provided by the DOE JGI website.

2.5.2 Full-length Gene Model Construction

RNA-seq data [Abbriano 2017, Smith et al. 2016] was visualized using the Integrated Genomics Viewer (IGV) [Robinson et al. 2011, Thorvaldsdottir et al. 2013]. Exon boundary coordinates were used to obtain genomic sequences on the DOE JGI website.

2.5.3 Predicted Protein Targeting Analysis

Open reading frames for the full-length gene models were obtained using the online ExPASy Translate tool (https://web.expasy.org/translate/). Because chloroplast-targeted diatom proteins must cross the ER membrane first, online programs SignalP 3.0 [Bendtsen et al. 2004] and SignalP 4.1 [Petersen et al. 2011] were used to predict ER targeting, and ChloroP 1.1 [Emanuelsson et al. 1999] was used to predict the presence of chloroplast transit domains, as previously discussed in Smith et al. [2012].

2.5.4 Additional Sequence-Based Analyses

Alignments and sequence identity analyses were performed with Clustal Omega (2.5.1). BCH sequences were obtained from the National Center for Biotechnology Information (NCBI) website (https://www.ncbi.nlm.nih.gov/) using GenBank accession numbers published by Tian and DellaPenna [2004]. BLASTp and tBLASTn programs on the NCBI website were used to search for proteins and translated nucleotide sequences with sequence identity to the C-terminal portion of Thaps3_263437 ("BCH"). Conserved protein domain and motif analyses were performed using the NCBI Conserved Domains Search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) [Marchler-Bauer et al. 2017], the ScanProsite tool (https://prosite.expasy.org/scanprosite/) [De

Castro et al. 2006], and InterPro (https://www.ebi.ac.uk/interpro/) [Finn et al. 2017]. Functional annotation predictions for the *T. pseudonana* carotenoid biosynthesis enzymes were obtained from the DOE JGI website as well as by protein sequence analysis using InterPro and Pfam (https://pfam.xfam.org/search/sequence). Structure-based functional annotation predictions were performed with Phyre2 [Kelley et al. 2015] (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index).

Additional diatom genomes were accessed at: http://genomes3.mcdb.ucla.edu/cgibin/hgGateway?hgsid=23003&clade=plant&org=Ahi+simulation +02&db=0 for *Cyclotella cryptica* (2011 Assembly), https://genome.jgi.doe.gov/Fracy1/Fracy1.home.html for *Fragilariopsis cylindrus*, https://genome.jgi.doe.gov/Psemu1/Psemu1.home.html for *Pseudonitzschia multiseries*, and https://genome.jgi.doe.gov/Thaoce1/Thaoce1.home.html for *Thalassiosira oceanica.*

Carotenoid biosynthesis genes in those genomes were found by using BLAT for *C. cryptica* and tBLASTn for the others, with predicted *T. pseudonana* protein sequences (2.3.4) and Phatr2_56492 (*P. tricornutum* ZEP3) as queries.

2.5.5 Genetic Manipulation

Cloning was carried out using MultiSite Gateway technology (Thermo Fisher Scientific, Waltham, MA, USA) as previously described [Shrestha and Hildebrand 2015]. VDL2 OE was driven by the *T. pseudonana* nitrate reductase (NR) promoter, LTL KD by the *T. pseudonana* ribosomal protein L41 promoter, and VDL1 and VDL2 KDs by the *T. pseudonana* acetyl CoA carboxylase (ACCase) promoter, with the corresponding terminators used respectively. As demonstrated previously [Shrestha et al. 2013], the rpL41 is the strongest promoter out of the three, and NR is the

weakest. The KD constructs comprised a nourseothricin resistance gene encoding nourseothricin N-acetyltransferase (NAT1) upstream 500-600 bp of antisense per gene of interest on the same transcript. Plasmid and primer details are available in **Appendix 2.E**. The constructs were transformed into WT *T. pseudonana* using tungsten microparticle bombardment with Bio-Rad PDS-1000/He, following procedures described by Davis et al. [2017], with the modification of incubating cells in dim light overnight following the bombardment. VDL2 OE was co-transformed with another plasmid carrying the NAT1 gene under the *T. pseudonana* ACCase promoter (received from N. Kroger, Germany). Genomic integration of the constructs was confirmed by PCR [Shrestha and Hildebrand 2015]. For VDL2 OE, RNA extraction, reverse transcription, qRT-PCR primer design, and qRT-PCR with normalization to the TATA box binding protein (Thaps3_264095) were performed as in Shrestha and Hildebrand [2015]. Four clones were chosen for further analysis **(Fig. S2.2)**.

2.5.6 Cultivation Conditions and Photopigment Analysis

T. pseudonana WT and transgenic cultures were cultivated at either 30 or 300 µmol photons m^{-2} sec^{-1} (natural white LED lighting, superbrightleds.com, NFLS-NW300X3-WHT-LC2), using a 12:12 h light:dark regime, at 18°C. 50 mL cultures in Erlenmeyer flasks were maintained in Artificial Sea Water (ASW) medium [Darley and Volcani 1969] with rapid stirring. Each experimental set included 2 WT and 4 transgenic cultures. After inoculation, the cultures were grown to 1-3x10^6 cells/mL, then allowed to adapt to the cultivation conditions by daily dilutions that maintained exponential growth with culture density under 2.5x10^6 cells/mL for a minimum of 2 weeks prior to sampling for photopigments. Cultures were rotated between stir plates each day to minimize any position-specific differences. Sampling was performed within the first two hours of the light period. Immediately prior to sampling, cultures were allowed to adapt to the light environment of the

biosafety cabinet for 45-90 min. 10-20 mL of $1x10^6$-$2.5x10^6$ cells/mL cultures were harvested, processed, and analyzed by HPLC as previously described [Kozlowski et al. 2011]. HPLC analysis employed an internal control of including the same sample in two different quantities to ensure precision in pigment quantification, which scaled accordingly. An online one-way ANOVA calculator accessed at https://www.socscistatistics.com/tests/anova/default2.aspx was used for statistical analysis of pigmentation differences between WT and transgenic lines.

2.6 ACKNOWLEDGEMENTS

Chapter 2, in full, is material currently being prepared for submission for publication. Gaidarenko, Olga; Mills, Dylan W.; Vernet, Maria; Hildebrand, Mark. "Novel *Thalassiosira pseudonana* violaxanthin de-epoxidase-like enzyme (VDL2) catalyzes fucoxanthin biosynthesis." Olga Gaidarenko was the principal researcher and author of this work. We thank Dr. Sarah R. Smith and Dr. Raffaela M. Abbriano for helpful discussions and data that made this work possible, as well as Dr. Roshan P. Shrestha for helpful discussions and providing genetic manipulation tools. Additionally, we thank Dr. Bradley Moore and Dr. Jonathan Chekan for providing valuable perspective and insights. We also thank Dr. James Golden for critically reading the chapter and providing helpful input. This work was supported by U.S. Dept. of Energy grant DE-FOA-0001471.

Table 2.1.

Model IDs of known carotenoid biosynthesis genes/enzymes and corresponding BLAST results.

Sequence Identity BLAST Results For:	Enzyme Name	*T. pseudonana*	*P. tricornutum*
Phytoene Synthase (PSY)	PSY1	Thaps3_268908	Phatr2_56481
	PSY2	Thaps3_263269	----------
Phytoene Desaturase (PDS)	PDS1	Thaps3_23291	Phatr2_45735
	PDS2	Thaps3_1383	Phatr2_55102
	----------	Thaps3_bd_1474	----------
ζ-Carotene Desaturase (ZDS)	ZDS	Thaps3_24832	Phatr2_53974
Carotene Cis-Trans Isomerase	----------	Thaps3_7094	Phatr2_45243
(Prolycopene Isomerase) (CRTISO)	----------	Thaps3_21900	Phatr2_9210
	----------	Thaps3_21847	Phatr2_54842
	----------	Thaps3_5221	Phatr2_51868
	----------	Thaps3_10233	Phatr2_54826
	----------	Thaps3_11636	Phatr2_54800
	----------	Thaps3_5859	Phatr2_42980
	----------	Thaps3_25361	----------
	----------	Thaps3_10254	----------
Lycopene β-cyclase (LCYB)	LCYB	Thaps3_270357	Phatr2_56484
β-Carotene Hydroxylase (BCH)	BCH	Thaps3_263437	----------
LUT-Like (Lutein Deficient-Like) (LTL)	LTL1	Thaps3_9541	Phatr2_50101
	LTL2	Thaps3_270336	Phatr2_26422
	----------	Thaps3_33926	Phatr2_34027
	----------	Thaps3_32491	Phatr2_33568
	----------	Thaps3_1549	Phatr2_6940
	----------	Thaps3_264647	Phatr2_46438
	----------	Thaps3_25944	Phatr2_31339
	----------	Thaps3_14875	Phatr2_47234
	----------	Thaps3_4027	Phatr2_37006
	----------	Thaps3_4026	Phatr2_43466
	----------	Thaps3_bd_518	Phatr2_43467
	----------	Thaps3_269400	Phatr2_43562
	----------	Thaps3_264325	Phatr2_50619
	----------	Thaps3_263399	Phatr2_43469
	----------	----------	Phatr2_32833
	----------	----------	Phatr2_43537

Table 2.1, continued.

Model IDs of known carotenoid biosynthesis genes/enzymes and corresponding BLAST results.

Zeaxanthin Epoxidase (ZEP)	ZEP1	Thaps3_270370	Phatr2_45845
	ZEP2	Thaps3_261390	Phatr2_56488
	ZEP3	----------	Phatr2_56492
	----------	Thaps3_1961	Phatr2_43425
	----------	Thaps3_6395	Phatr2_47925
	----------	Thaps3_20663	Phatr2_45936
	----------	Thaps3_22671	----------
Violaxanthin De- Epoxidase (VDE),	VDE	Thaps3_7677	Phatr2_44635
Like (VDL), VDE-Related (VDR)	VDL1	Thaps3_22076	Phatr2_46155
	VDL2	Thaps3_11707	Phatr2_45846
	VDR	Thaps3_270211	Phatr2_56450
	----------	----------	Phatr2_bd_1281

Table 2.2.
Targeting predictions for known carotenoid biosynthesis enzymes and candidates.
*ChloroP score very close to the cutoff (0.500), chloroplast targeting possible.

Predicted Targeting	Gene/Protein ID	Name	Notes
Chloroplast	Thaps3_268908	PSY1	
	Thaps3_23291	PDS1	
	Thaps3_1383	PDS2	
	Thaps3_24832	ZDS	
	Thaps3_270357	LCY-B	
	Thaps3_9541	LTL1	
	Thaps3_270370	ZEP1	
	Thaps3_261390	ZEP2	
	Thaps3_7677	VDE	
	Thaps3_22076	VDL1	
	Thaps3_11707	VDL2	
	Thaps3_270211	VDR	
	Thaps3_21900		Found in CRTISO BLAST search.
	Thaps3_10233		Found in CRTISO BLAST search.
Endoplasmic	Thaps3_263269	PSY2*	ChloroP Score = 0.499, close to the 0.500 cutoff.
Reticulum (ER)	Thaps3_270336	LTL2*	ChloroP Score = 0.497, close to the 0.500 cutoff.
	Thaps3_14875		Clear ER targeting, possible signal anchor. Clear lack of predicted chloroplast targeting.
	Thaps3_264647		Clear ER targeting, possible signal anchor. Clear lack of predicted chloroplast targeting.
	Thaps3_5221		Clear ER targeting, possible signal anchor. Clear lack of predicted chloroplast targeting.
Other	Thaps3_bd_1474	PDS3	Missing N-terminus, prediction unavailable. Appears identical to PDS1 (2.5.1).
	Thaps3_263437	BCH	Clear lack of ER or chloroplast.
	Thaps3_4026		Clear lack of ER or chloroplast.
	Thaps3_25361		Clear chloroplast, but not ER.
	Thaps3_33926		Clear chloroplast, but not ER.
	Thaps3_1549		Clear chloroplast, but not ER.
	Thaps3_6395		Clear chloroplast, but not ER.
	Thaps3_bd_518		Unclear, RNA-seq data unavailable to confirm gene model.

Table 2.3.
Partial sequence identity matches for the Thaps3_263437 C-terminal, non-BCH-like peptide.

Portion of C-terminal part	BLAST Result Type	Sequence Identity	Organism
81-88%	Unidentified or hypothetical proteins	28-39%	*Chrysochromulina* sp. CCMP291 (haptophyte) *Aurantiochytrium* sp. FCC1311 (labyrinthulomycete) *Oikopleura dioica* (tunicate)
26-37%, C-terminal	Translated nucleotide sequences	41-43%	*Oikopleura dioica* (tunicate) *Homo sapiens* (primate) *Pan troglodytes* (primate)
17%, middle	Translated nucleotide sequence	38%	*Arachis hypogaea* (green plant)

Table 2.4.
Sequence-based functional annotation of *T. pseudonana* carotenoid biosynthesis enzymes.

Name/ID	Gene Ontology (GO)	EuKaryotic Orthologous Group (KOG)	Protein Family (Pfam)	InterPro Family
PSY1 Thaps3_268908	GO:0009058 biosynthetic process GO:0016740 transferase activity	KOG1459 Squalene synthetase	PF00494 Squalene/phytoene synthase	IPR002060 Squalene/phytoene synthase IPR008949 Isoprenoid synthase domain superfamily
PSY2 Thaps3_263269	GO:0003824 catalytic activity GO:0008299 isoprenoid biosynthetic process GO:0009058 biosynthetic process GO:0009507 chloroplast GO:0016117 carotenoid biosynthetic process GO:0016740 transferase activity	KOG1459 Squalene synthetase	PF00494 Squalene/phytoene synthase	IPR002060 Squalene/phytoene synthase IPR008949 Isoprenoid synthase domain superfamily
PDS1 Thaps3_23291	GO:0004497 monooxygenase activity GO:0009507 chloroplast GO:0016117 carotenoid biosynthetic process GO:0016491 oxidoreductase activity	KOG0029 Amine oxidase	PF01593 Flavin-containing amine oxidase	IPR014102 Phytoene desaturase IPR036188 FAD/NAD(P)-binding domain superfamily IPR002937 Amine oxidase

Table 2.4, continued.
Sequence-based functional annotation of *T. pseudonana* carotenoid biosynthesis enzymes.

Thaps3_bd_1474	GO:0016491 oxidoreductase activity	KOG0029 Amine oxidase	PF01593 Flavin-containing amine oxidase	IPR036188 FAD/NAD(P)-binding domain superfamily IPR002937 Amine oxidase
PDS2 Thaps3_1383	GO:0009507 chloroplast	KOG0029 Amine oxidase	PF01593 Flavin-containing amine oxidase	IPR014102 Phytoene desaturase
	GO:0016117 carotenoid biosynthetic process GO:0016120 carotene biosynthetic process			IPR036188 FAD/NAD(P)-binding domain superfamily
	GO:0016166 phytoene dehydrogenase activity GO:0016491 oxidoreductase activity			IPR002937 Amine oxidase
ZDS Thaps3_24832	GO:0016117 carotenoid biosynthetic process	KOG0029 Amine oxidase	PF01593 Flavin-containing amine oxidase	IPR014103 Zeta-carotene desaturase
	GO:0016491 oxidoreductase activity GO:0016719 carotene 7,8-desaturase activity			IPR036188 FAD/NAD(P)-binding domain superfamily IPR002937 Amine oxidase
CRTISO Thaps3_21900	GO:0016117 carotenoid biosynthetic process	KOG4254 Phytoene desaturase	PF13450 NAD(P)-binding Rossmann-like domain	IPR036188 FAD/NAD(P)-binding domain superfamily
	GO:0016705 oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen GO:0050660 flavin adenine dinucleotide binding			
Thaps3_10233	GO:0015036 disulfide oxidoreductase activity	KOG4254 Phytoene desaturase	PF13450 NAD(P)-binding Rossmann-	IPR036188 FAD/NAD(P)-binding domain superfamily

Table 2.4, continued.
Sequence-based functional annotation of *T. pseudonana* carotenoid biosynthesis enzymes.

	GO:0016117 carotenoid biosynthetic process GO:0016491 oxidoreductase activity GO:0016705 oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen GO:0050660 flavin adenine dinucleotide binding	PF01593 Flavin-containing amine oxidase		IPR002937 Amine oxidase
LCY-B Thaps3_270357	GO:0008152 metabolic process GO:0009507 chloroplast GO:0016117 carotenoid biosynthetic process GO:0016491 oxidoreductase activity GO:0016705 oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	N/A	PF05834 Lycopene cyclase	IPR010108 Lycopene cyclase, beta/epsilon IPR036188 FAD/NAD(P)-binding domain superfamily
LTL1 Thaps3_9541	GO:0004497 monooxygenase activity	N/A	PF00067 Cytochrome P450	IPR001128 Cytochrome P450 IPR002401 Cytochrome P450, E-class, group I
LTL2 Thaps3_270336	GO:0004497 monooxygenase activity GO:0005506 iron ion binding	N/A	PF00067 Cytochrome P450	IPR001128 Cytochrome P450 IPR002401 Cytochrome P450, E-class, group I

Table 2.4, continued.
Sequence-based functional annotation of *T. pseudonana* carotenoid biosynthesis enzymes.

	GO:0020037 heme binding			IPR036396 Cytochrome P450 superfamily
ZEP1 Thaps3_270370	GO:0004497 monooxygenase activity	KOG2614 Kynurenine 3-monooxygenase and related flavoprotein monooxygenases	PF01494 FAD binding domain	IPR036188 FAD/NAD(P)-binding domain superfamily
	GO:0006118 obsolete electron transport GO:0006725 cellular aromatic compound metabolic process GO:0008152 metabolic process GO:0009507 chloroplast GO:0016491 oxidoreductase activity			IPR002938 FAD-binding domain
ZEP2 Thaps3_261390	GO:0004497 monooxygenase activity	KOG2614 Kynurenine 3-monooxygenase and related flavoprotein monooxygenases	PF01494 FAD binding domain	IPR036188 FAD/NAD(P)-binding domain superfamily
	GO:0006725 cellular aromatic compound metabolic process GO:0008152 metabolic process GO:0016123 xanthophyll biosynthetic process GO:0016491 oxidoreductase activity			IPR002938 FAD-binding domain IPR003953 FAD-dependent oxidoreductase 2, FAD binding domain
VDE	GO:0055114 oxidation-	N/A	PF07137 VDE	IPR010788 VDE lipocalin domain

Table 2.4, continued.
Sequence-based functional annotation of *T. pseudonana* carotenoid biosynthesis enzymes.

	GO:0046422 violaxanthin de-epoxidase activity GO:0009507 chloroplast			IPR012674 Calycin
VDL1 Thaps3_22076	GO:0055114 oxidation-reduction process GO:0046422 violaxanthin de-epoxidase activity GO:0009507 chloroplast	N/A	PF07137 VDE lipocalin domain	IPR010788 VDE lipocalin domain IPR012674 Calycin
VDL2 Thaps3_11707	GO:0055114 oxidation-reduction process GO:0046422 violaxanthin de-epoxidase activity GO:0009507 chloroplast	KOG4157 β-1,6-N-acetylglucosaminyl-transferase, contains WSC domain	PF07137 VDE lipocalin domain	IPR010788 VDE lipocalin domain IPR012674 Calycin
VDR Thaps3_270211	N/A	N/A	N/A	IPR012674 Calycin

Table 2.5.
Carotenoid biosynthesis genes in currently available diatom genomes.

	T. pseudonana	*T. oceanica*	*C. cryptica*	*P. tricornutum*	*F. cylindrus*	*P. multiseries*
PSY1	Thaps3_268908	scaffold_1529: 4644-6041	g3020512_08749	Phatr2_56481	Fracy1_264173	Psemu1_284362
PSY2	Thaps3_263269	Thaoce1_76633	g1711422_03357	----------	Fracy1_233859	Psemu1_21973
	----------	----------	----------	----------	scaffold_81: 89953-89306	Psemu1_252566
PDS1	Thaps3_23291	Thaoce1_74122	g3033530_08435	Phatr2_45735	scaffold_15: 196010-194589	Psemu1_296799
	Thaps3_bd_1474?	----------	g3033529_10333	----------	----------	----------
	----------	----------	g2978166_08271	----------	----------	----------
PDS2	Thaps3_1383	Thaoce1_69579	g3041196_11323	Phatr2_55102	scaffold_41: 345681-344246	Psemu1_184015
	----------	----------	----------	----------	Fracy1_260963	----------
ZDS	Thaps3_24832	Thaoce1_79521	g3029884_03203	Phatr2_53974	Fracy1_291551	Psemu1_201585
	----------	----------	----------	----------	scaffold_28: 522744-521104	----------
	----------	----------	----------	----------	Fracy1_231297	----------
CRTISO	Thaps3_21900	scaffold_8852: 1169-2900	g3031469_09996	Phatr2_45243	scaffold_15: 1032105-1030436	Psemu1_296898
	----------	----------	----------	----------	Fracy1_225509	----------
CRTISO or Fx Synth.	Thaps3_10233	Thaoce1_83016+83017	g3032342_19300	Phatr2_9210	Fracy1_274697	Psemu1_238122
LCY-B	Thaps3_270357	Thaoce1_81190	g2984704_14008	Phatr2_56484	Fracy1_183412	Psemu1_181616
	----------	----------	----------	----------	Fracy1_190208	----------
LTL1	Thaps3_9541	Thaoce1_93110+93111	g3003680_30430	Phatr2_50101	Fracy1_261383	Psemu1_242952
LTL2	Thaps3_270336	Thaoce1_94290	g3010342_02309	Phatr2_26422	scaffold_46: 478885-477109	Psemu1_249715

Table 2.5, continued.
Carotenoid biosynthesis genes in currently available diatom genomes.

	----------	----------	g3029177_02308	----------	----------	----------
ZEP1	Thaps3_270370	scaffold_19419:2-1117	g2969398_36291	Phatr2_45845	Fracy1_232148	Psemu1_318239
	----------	Thaoce1_86217	----------	----------	----------	----------
ZEP2	Thaps3_261390	Thaoce1_91593	g3012830_03620	Phatr2_56488	Fracy1_208380	Psemu1_249822
ZEP3	----------	----------	g3012829_03620	Phatr2_56492	Fracy1_260743	Psemu1_321577
VDE	Thaps3_7677	scaffold_16: 15273-16483	g3000467_15552	Phatr2_44635	Fracy1_267113	Psemu1_282298
VDL1	Thaps3_22076	Thaoce1_92196	g3024877_16840	Phatr2_46155	Fracy1_212709	Psemu1_312307
VDL2	Thaps3_11707	Thaoce1_75745	g2981206_02202	Phatr2_45846	Fracy1_291552	Psemu1_252938
	----------	----------	g2981205_02344	Phatr2_bd_1281	----------	----------
	----------	----------	g2981208_01259	----------	----------	----------
	----------	----------	g2981207_01260	----------	----------	----------
VDR	Thaps3_270211	Thaoc1_72825+72826	g2981650_01688	Phatr2_56450	scaffold_62: 310847-309744	Psemu1_170800
	----------	----------	g3028784_03450	----------	Fracy1_269417	----------
	----------	----------	g2990797_01166	----------	----------	----------
	----------	----------	g3022545_01165	----------	----------	----------

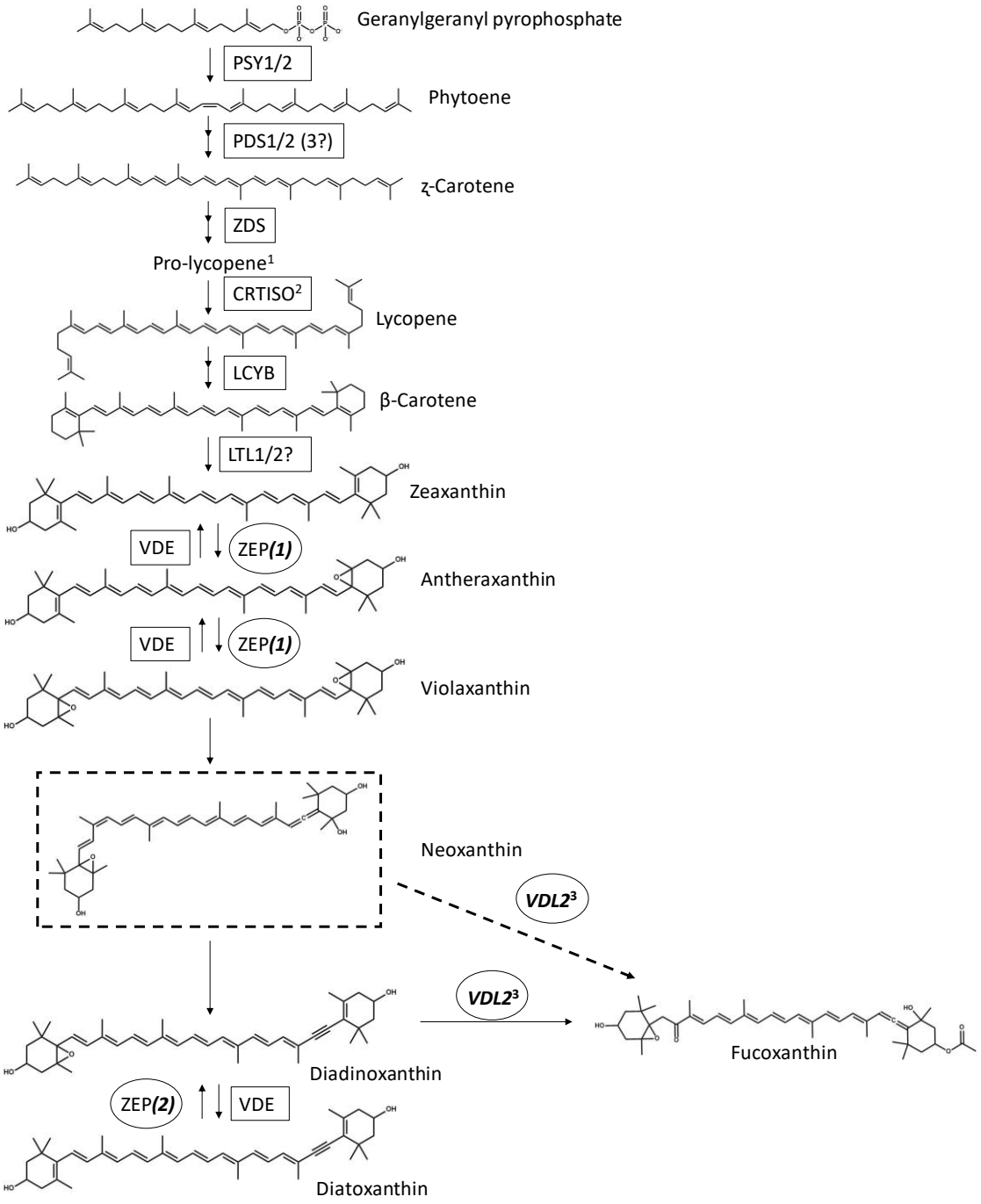


Figure. 2.1.

Putative carotenoid biosynthesis pathway in *T. pseudonana.* Enzymes assigned to steps based on this study are depicted in ovals in bold italics. [1]Pro-lycopene is a stereoisomer of lycopene. [2]Thaps3_21900 is hypothesized to catalyze this step. Thaps3_10233 might also be involved.[3]Thaps3_10233 might also be involved. PSY = phytoene synthase, PDS = phytoene desaturase, ZDS = ζ-carotene desaturase, CRTISO = carotene cis/trans isomerase, prolycopene isomerase, LCYB = lycopene cyclase b, LTL = LUT-like, ZEP = zeaxanthin epoxidase, VDE = violaxanthin de-epoxidase, VDL = VDE-like.

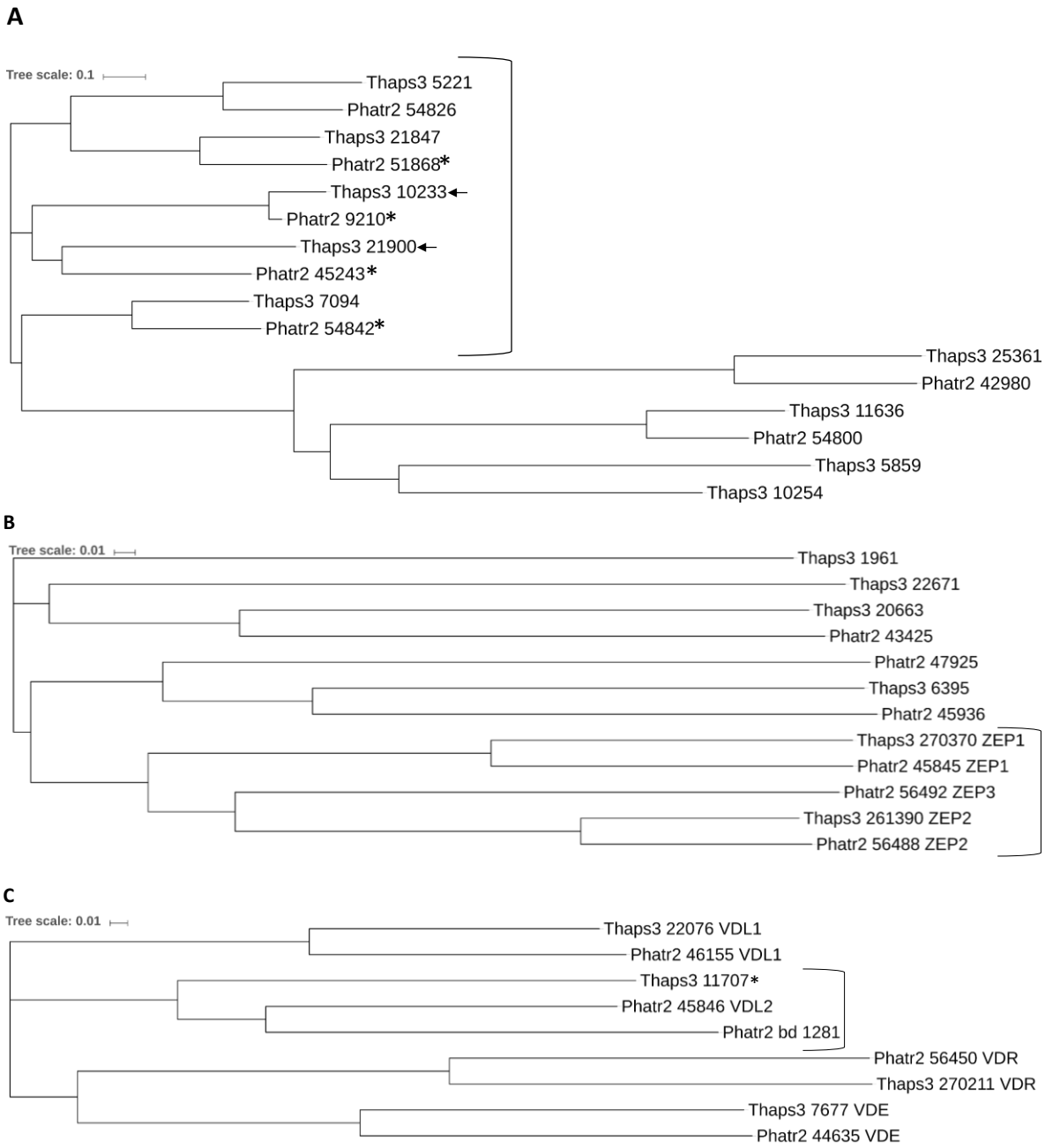


Figure. 2.2.

Sequence identity-based phylogenetic trees. **A.** CRTISO candidates and related proteins. The bracketed group includes those previously identified as CRTISO candidates by Dambeck et al. [2012], indicated with an asterisk (*). Candidates hypothesized to be involved in carotenoid biosynthesis based on transcriptomic analysis and predicted chloroplast targeting are indicated with an arrow (←). **B.** ZEPs (bracketed) and related proteins. **C.** Known VDEs, VDLs, VDRs, and related proteins. Group including the *P. tricornutum* VDL2 is bracketed. The novel *T. pseudonana* protein Thaps3_11707, hereafter designated VDL2, is indicated with an asterisk (*).

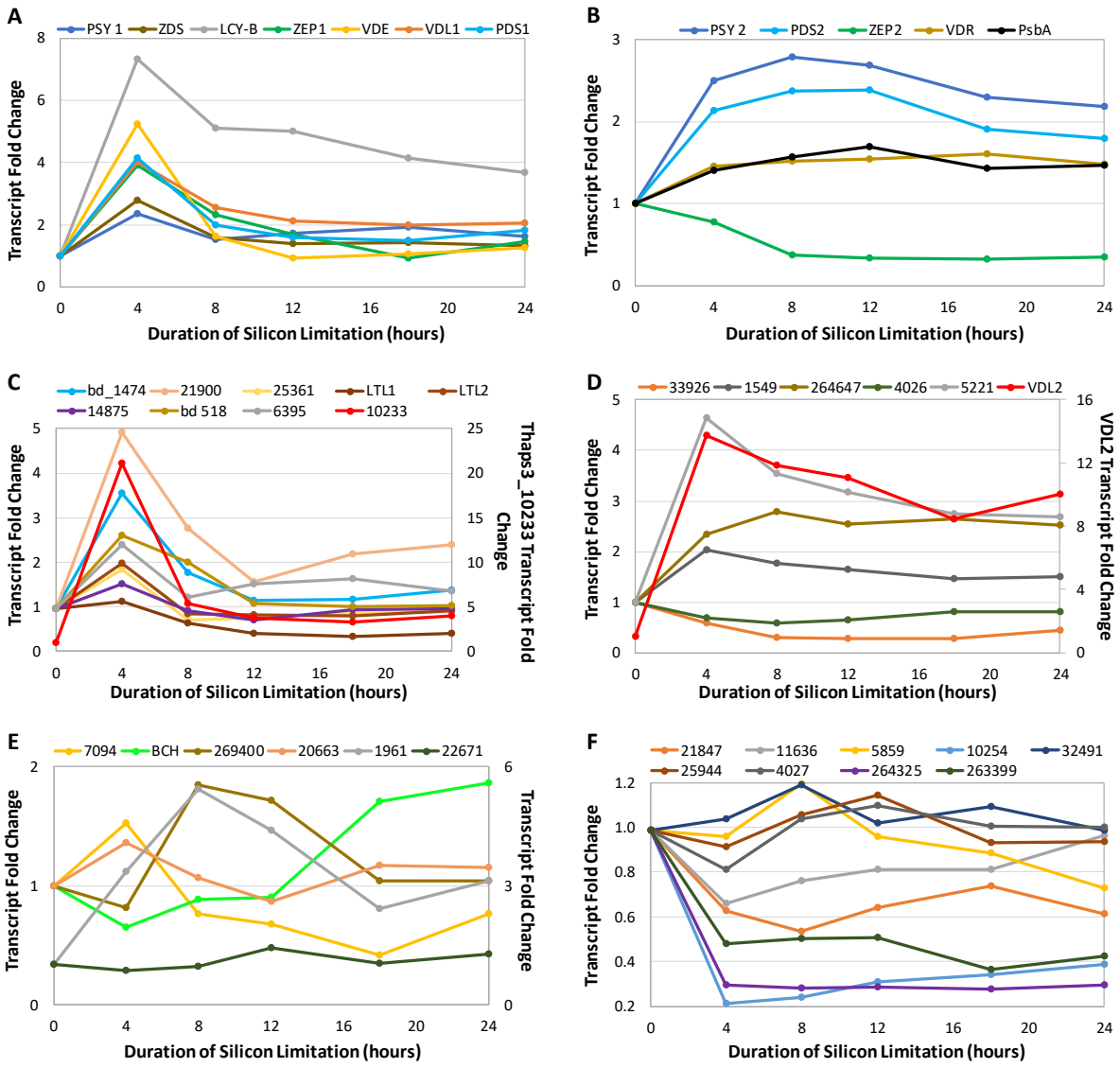


Figure. 2.3.

Candidate gene silicon starvation microarray expression patterns [Smith et al. 2016].

A, B. Genes known to be involved in carotenoid biosynthesis (except PsbA, which encodes the photosystem II D1 protein);

C, D. Genes with expression patters similar to those in A and B, chosen for further study.

E, F. Genes with expression patterns different from those in A and B, not chosen for further study.

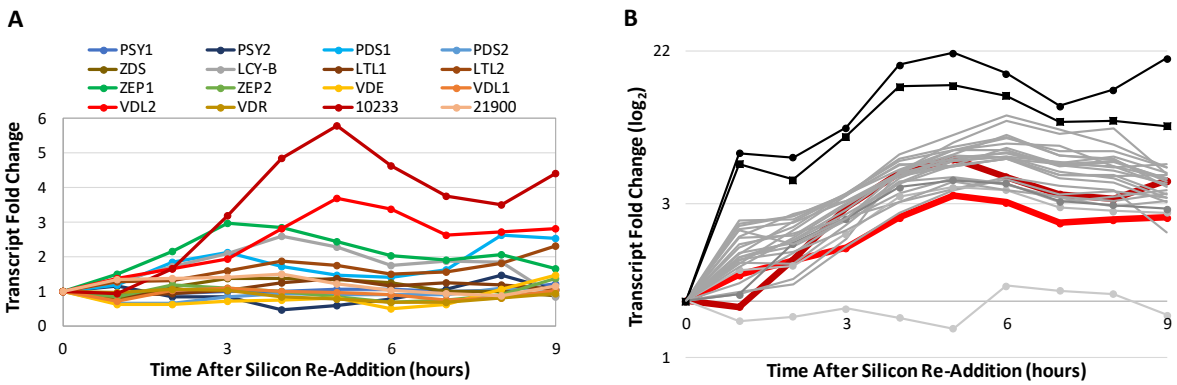


Figure. 2.4.

RNA-seq gene expression patterns [Abbriano 2017].

A. Carotenoid biosynthesis genes; **B.** VDL2 (bright red), Thaps3_12033 (dark red), and photoantenna protein genes: Thaps3_32723 (light grey, circles), Thaps3_38667 (black, circles), Thaps3_42962 (black, squares), Thaps3_2601, 2845, 3815, 5174, 6139, 7916, 10219, 29375, 30385, 31749, 31983, 33018, 33131, 33606, 34276, 36081, 38122, 39813, 40747, 262313, 262332, 268127, 270092 (dark grey).

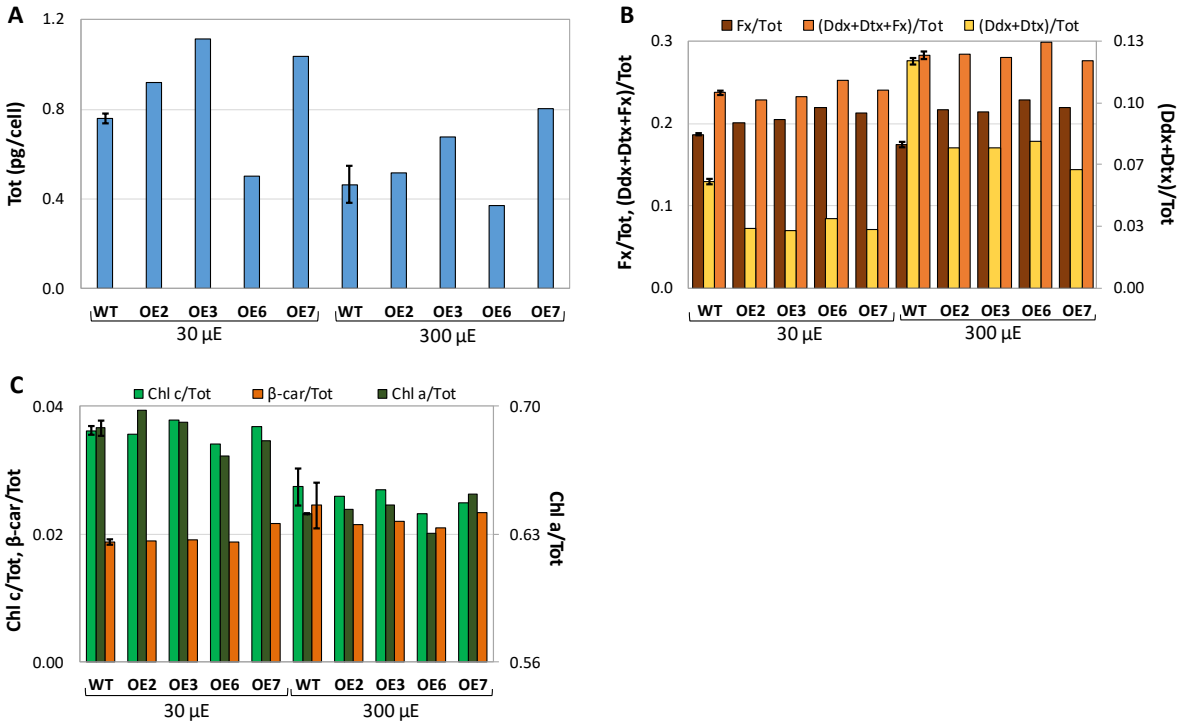


Figure. 2.5.

HPLC-based pigment analysis of wild-type (WT) vs. VDL2 overexpression (OE) lines cultured at 30 µmol photons m^{-2} sec^{-1} (µE) and 300 µE. WT data is an average of two independent cultures. **A**. Total cellular photopigments (Tot); **B.** Fx/Tot, (Ddx+Dtx)/Tot, (Ddx+Dtx+Fx)/Tot; **C.** Chl a/Tot, Chl c/Tot, β-car/Tot.

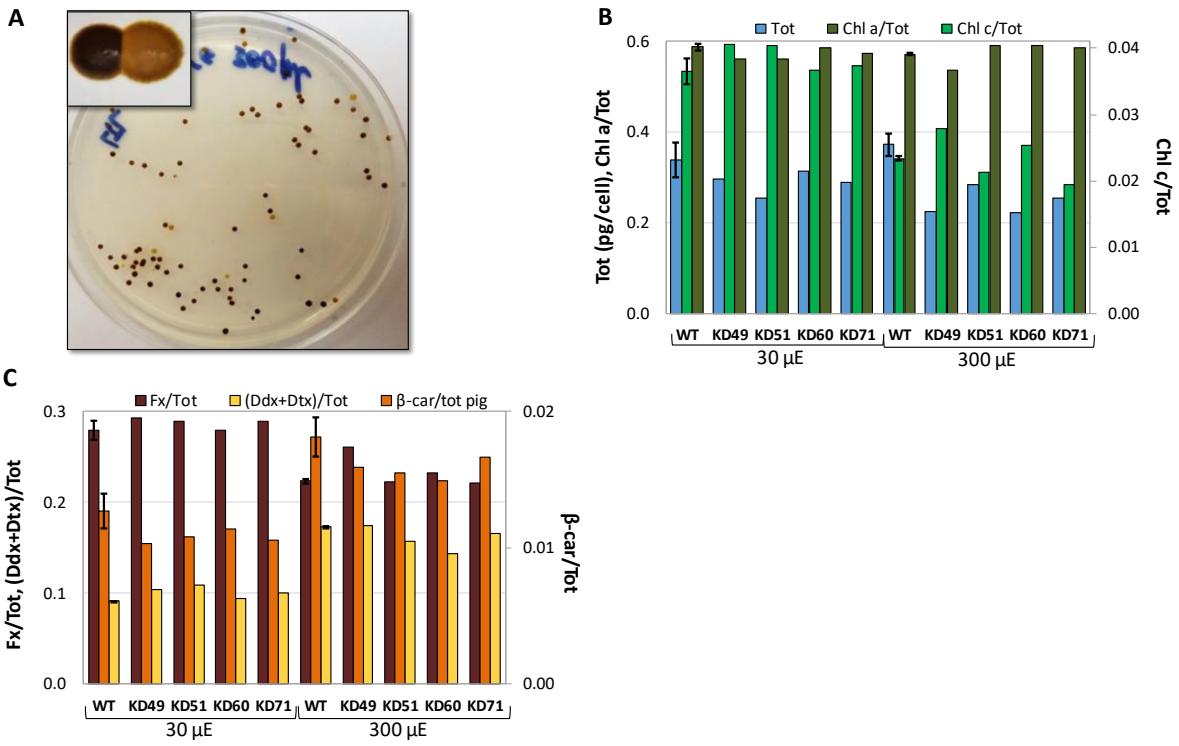


Figure. 2.6.

A. Lighter-pigmented LTL knockdown (KD) colonies among those with wild-type (WT)-equivalent pigmentation; HPLC-based pigment analysis of LTL KD clones cultured at 30 µmol photons m^{-2} sec^{-1} (µE) and 300 µE. WT data is an average of two independent cultures. **B.** Tot, Chl a/Tot, Chl c/Tot; **C.** Fx/Tot, (Ddx+Dtx)/Tot, β-car/Tot.

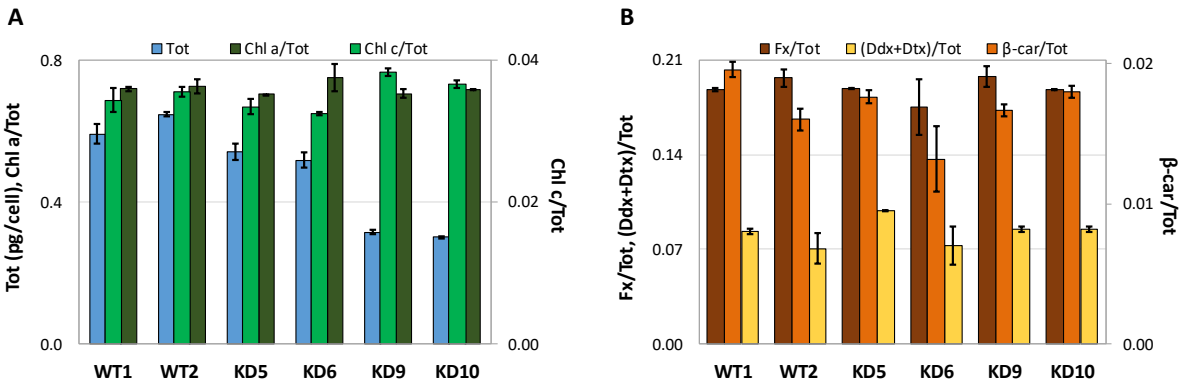


Figure. 2.7.
HPLC-based pigment analysis of wild-type (WT) vs. VDL1 knockdown (KD) clones cultured at 300 µmol photons m^{-2} sec^{-1}. Each data point is an average of two samples taken from the same culture. **A.** Tot, Chl a/Tot, Chl c/Tot; **B.** Fx/Tot, (Ddx+Dtx)/Tot, β-car/Tot.

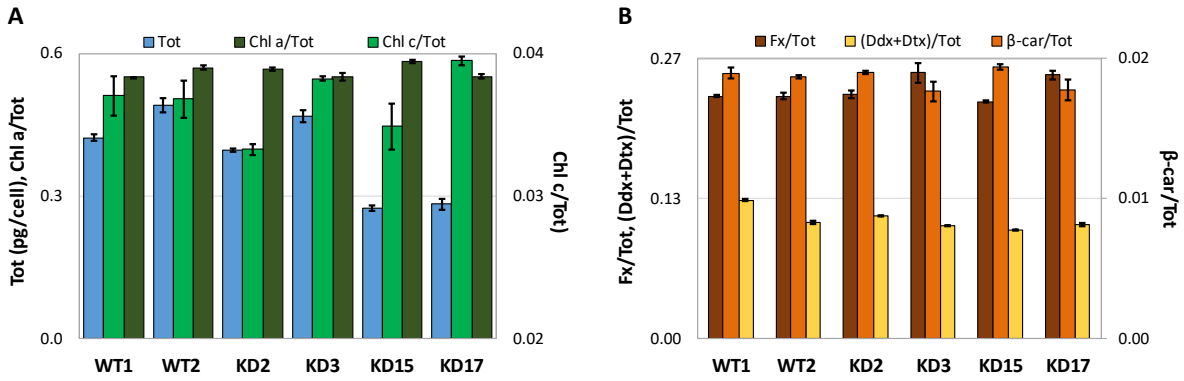


Figure. 2.8.
HPLC-based pigment analysis of wild-type (WT) vs. VDL2 knockdown (KD) clones cultured at 300 µmol $m^{-2}sec^{-1}$. Each data point is an average of two samples taken from the same culture. **A.** Tot, Chl a/Tot, Chl c/Tot; **B.** Fx/Tot, (Ddx+Dtx)/Tot, β-car/Tot.

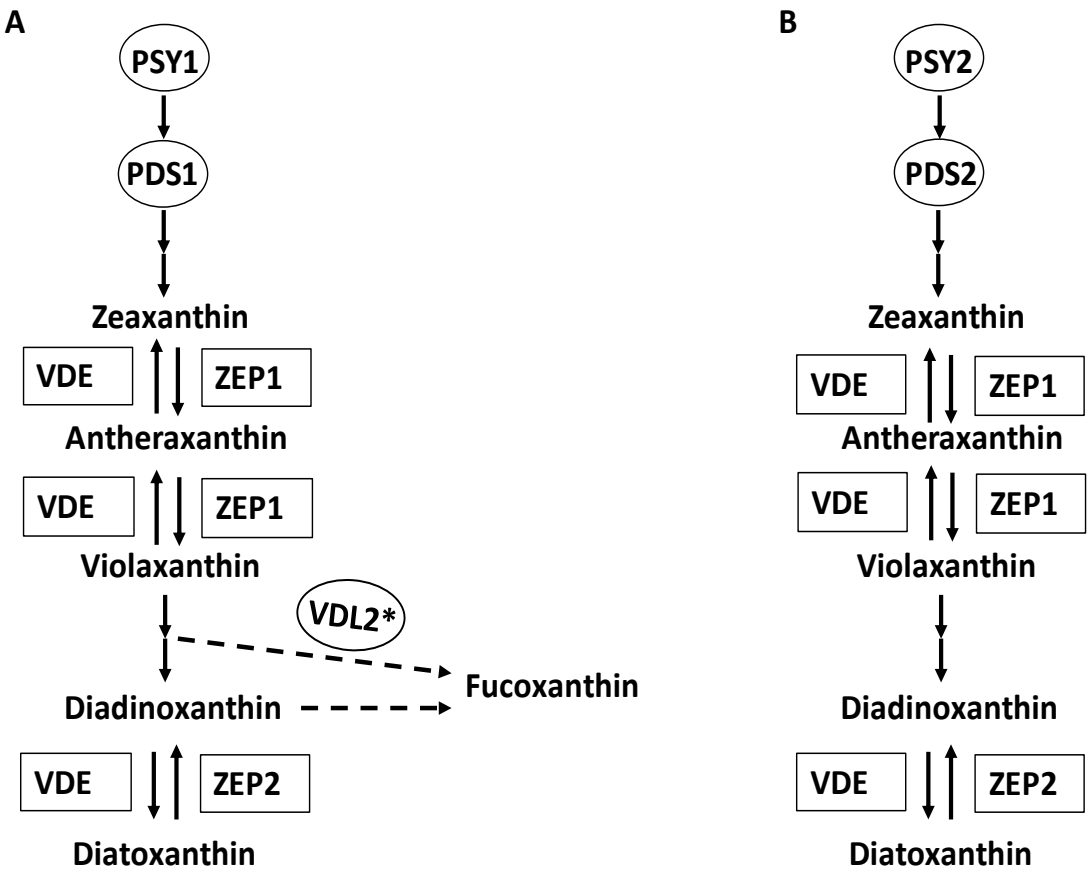


Figure. 2.9.
Model of differential carotenoid biosynthesis regulation in *T. pseudonana.* Differentially utilized enzymes are designated by ovals. **A.** During chloroplast replication: some flux is directed towards Fx biosynthesis to populate photoantenna proteins. *Thaps3_10233 might also be involved. **B.** During an increase in illumination: Ddx+Dtx accumulate, mostly in the lipid shield around photoantennae.

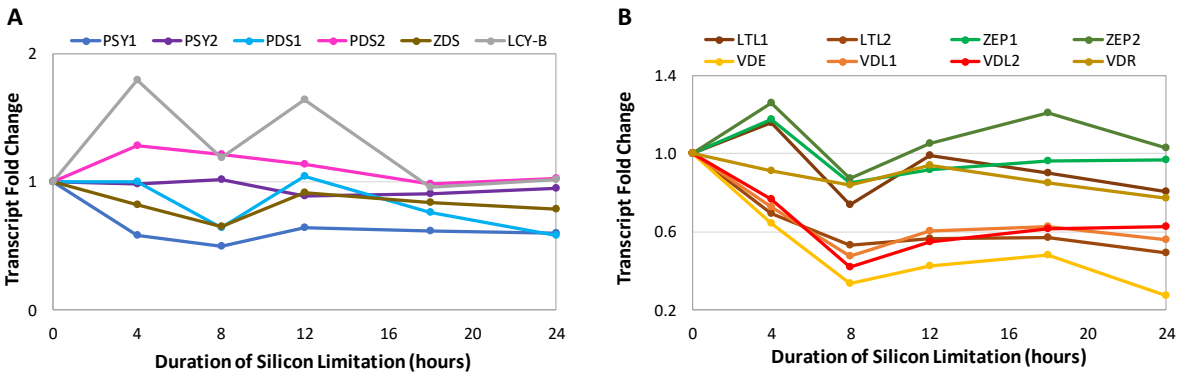


Figure. S2.1.
RNA-seq carotenoid biosynthesis gene expression patterns during silicon starvation [Smith et al. 2016]. **A.** Pre-β-car part of the carotenoid biosynthesis pathway; **B.** Post-β-car part of the carotenoid biosynthesis pathway.

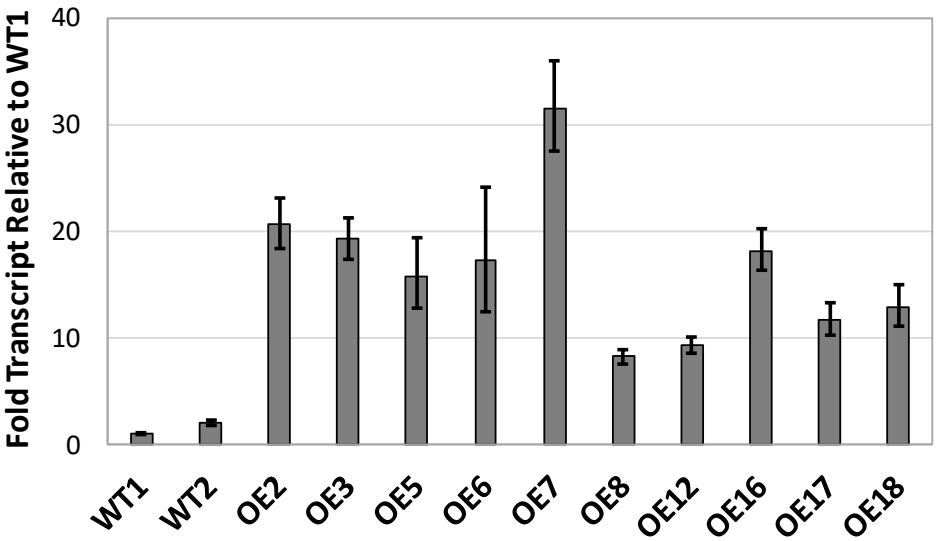


Figure. S2.2.
qRT-PCR screen for VDL2 overexpression (OE) compared to wild-type (WT) transcript levels. Clones 2, 3, 6, 7 were chosen for analysis. Each data point represents an average of two wells.

2.7 REFERENCES

Abbriano, R. (2017) Insights into the molecular regulation of growth and carbon flux in marine diatoms. Diss. University of California, San Diego.

Armbrust E. V., Berhes J. A., Bowler C., Green B. R., Martinez D., Putnam N. H., Zhou S., Allen A. E., Apt K. E., Bechner M., Brzezinski M. A., Chaal B. K., Ciovitti A., Davis A. K., Demarest M. S., Detter J. C., Glavina T., Goodstein D., Hazi M. Z., Hellsten U., Hildebrand M., Jenkins B. D., Jurka J., Kapitonov V. V., Kroger N., Lau W. W., Lane T. W., Larimer F. W., Lippmeier J. C., Lucas S., Medina M., Montsat A., Obornik M., Parker M. S., Palenik B., Pazour J. G., Richardson P. M., Rynearson T. A., Saito M. A., Schwartz D. C., Thamatrakoln K., Valentin K., Vardi A., Wilkerson F. P., Rokhsar D. S. (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science 306(5693): 79-86.

Bendtsen J. D., Nielsen H., von Heijne G., Brunak S. (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340: 783-795.

Bertrand M. (2010) Carotenoid biosynthesis in diatoms. Photosynth Res 106: 89-102.

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., VerretF., Berges J. A., Brownlee C., CadoretJ., Chiovitti A., Choi C. J., Coesel S., De Martino A., Detter J. C., Durkin C., Falciatore A., FournetJ., Haruta M., Huysman M. J. J., Jenkins B. D., Jiroutova K., Jorgensen R. E., Joubert Y., Kaplan A., Kroger N., Kroth P. G., La Rouche J., Lindquist E., Lommer M., Martin-Jezequel V., Lopez P. J., Lucas S., Mangogna M., McGinnis K., Medlin L. K., Montsant A., Oudot-Le Secq M., Napoli C., Obornik M., Parker M. S., Petit J., Porcel B. M., Poulsen N., Robison M., Rychlewski L., Rynearson T. A., Schmutz J., Shapiro H., Siaut M., Stanley M., Sussman M. R., Taylor A. R., Vardi A., von Dassow P., Vyverman W., Willis A., Wyrwicz L. S., Roksrar D. S., Weissenbach J., Armbrust E. V., Green B. R., Van de Peer Y., Grigoriev I. V. (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. Nature 456: 239-244.

BrunetC., Chandrasekaran R., Barra L., Giovagnetti V., Corato F., Ruban A. V. (2014) Spectral radiation dependent photoprotective mechanism in the diatom *Pseudo-nitzschia multistriata*. PLoS One 9(1): e87015.

Chamovitz D., Sandmann G., Hirschberg J. (1993) Molecular and biochemical characterization of herbicide-resistant mutants of cyanobacteria reveals that phytoene desaturation is a rate-limiting step in carotenoid biosynthesis. J Biol Chem 268(23): 17348-17353.

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(8): e2896.

Dambek M., Eilers U., Breitenbach J., Steiger S., Buchel C., Sandmann G. (2012) Biosynthesis of fucoxanthin and diadinoxanthin and function of initial pathway genes in *Phaeodactylum tricornutum*. J Exp Bot 63(15): 5607-5612.

Darley W., Volcani B. (1969) Role of silicon in diatom metabolism: a silicon requirement for deoxyribonucleic acid synthesis in the diatom *Cylindrotheca fusiformis* Reimann and Lewin. Exp Cell Res 58: 334–42.

Davis A., Abbriano R., Smith S. R., Hildebrand M. (2017) Clarification of photorespiratory processes and the role of malic enzyme in diatoms. Protist 168(1): 134-153.

De Castro E., Sigrist C. J. A., Gattiker A., Bulliard V., Langendijk-Genevaux P. S., Gasteiger E., Bairock A., Hulo N. (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res 34(Web Server Issue): W362-W365.

Domingues N., Matos A. R., da Silva J. M., Cartaxana P. (2012) Response of the diatom *Phaeodactylum tricornutum* to photooxidative stress resulting from high light exposure. PLoS One 7(6): e38162.

Eilers U., Bikoulis A., Breitenbach J., Buchel C., Sandmann G. (2016) Limitations in the biosynthesis of fucoxanthin as targets for genetic engineering in *Phaeodactylum tricornutum*. J Appl Phycol 28(1): 123-129.

Emanuelsson O., Nielsen H., von Heijne G. (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci 8(5): 978-984.

Finn R. D., Attwood T. K., Babbitt P. C., Bateman A., Bork P., Bridge A. J., Chang H., Dosztányi Z., El-Gebali S., Fraser M., Gough J., Haft D., Holliday G. L., Huang H., Huang X., Letunic I., Lopez R., Lu S., Marchler-Bauer A., Mi H., Mistry J., Natale D. A., Necci M., Nuka G., Orengo G. A., Park Y., Pesseat S., Piovesan D., Potter S. C., Rawlings N. D., Redaschi N., Richardson L., Rivoire C., Sangrador-Vegas A., Sigrist C., Sillitoe I., Smithers B., Squizzato S., Sutton G., Thanki N., Thomas P. D., Tosatto S. C. E., Wu C. H., Xenarios I., Yeh L., Young S., Mitchell A. L. (2017) InterPro in 2017 – beyond protein family and domain annotations. Nucleic Acids Res 45(Database Issue): D190-D199.

Goericke R., Welschmeyer N. A. (1992) Pigment turnover in the marine diatom *Thalassiosira weisflogii.* II The $^{14}CO_2$-labeling kinetics of carotenoids. J Phycol 28: 507-517.

Goss R., Pinto E. A., Wilhelm C., Richter M. (2006) The importance of a highly active and ΔpH-regulated diatoxanthin epoxidase for the regulation of the PSII antenna function in diadinoxanthin cycle containing algae. J Plant Physiol 163(10): 1008-1021.

Hildebrand M., Davis A. K., Smith S. R., Traller J. C., Abbriano R. (2012) The place of diatoms in the biofuels industry. Biofuels 3(2): 221-240.

Jakob T., Goss R., Wilhelm C. (2001) Unusual pH-dependence of diadinoxanthin de-epoxidase activation causes chlororespiratory induced accumulation of diatoxanthin in the diatom *Phaeodactylum tricornutum*. J Plant Physiol 158: 383-390.

Kadono T., Kira N., Suzuki K., Iwata O., Ohama T., Okada S., Nishimura T., Akakabe M., Tsuda M., Adachi M. (2015) Effect of an introduced phytoene synthase gene expression on carotenoid biosynthesis in the marine diatom *Phaeodactylum tricornutum*. Mar Drugs 13(8): 5334-5357.

Kaur S., Spillane C. (2015) Reduction in carotenoid levels in the marine diatom *Phaeodactylum tricornutum* by artificial microRNAs targeted against the endogenous *phytoene synthase* gene. Mar Biotechnol 17(1): 1-7.

Kelley L. A., Mezulis S., Yates C. M., Wass M. N., Sternberg M. J. E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10: 845-858.

Kirst H., Garcia-Cerdan J. G., Zurbiggen A., Ruehle T., Melis A. (2012)Truncated photosystem chlorophyll antenna size in the green microalga *Chlamydomonas reinhardtii* upon deletion of the *TLA3-CpSRP43* gene. Plant Physiol 160(4): 2251-2260.

Kozlowski W. A., Deutschman D., Garibotti I., Trees C., Vernet M., An evaluation of the application of CHEMTAX to Antarctic coastal pigment data. Deep-Sea Res. Pt I 58 (2011) 350-364.

Lavaud J., Lepetit B. (2012) An explanation for the inter-species variability of the photoprotective non-photochemical chlorophyll fluorescence quenching in diatoms. Biochim Biophys Acta 1827(3): 294-302.

Lavaud J., Materna A. C., Sturm S., Vugrinec S., Kroth P. G. (2012) Silencing of the violaxanthin de-epoxidase gene in the diatom *Phaeodactylum tricornutum* reduces diatoxanthin synthesis and non-photochemical quenching. PLoS One 7(5): e36806.

Lavaud J., Rousseau B., Etienne A. (2004) General features of photoprotection by energy dissipation in planktonic diatoms. J Phycol 40: 130-137.

Lepetit B., Goss R., Jakob T., Wilhelm C. (2012) Molecular dynamics of the diatom thylakoid membrane under different light conditions. Photosynth Res 111(1-2): 245-257.

Lepetit B., Volke D., Gilbert M., Wilhelm C., Goss R. (2010) Evidence for existence of one antenna-associated, lipid-dissolved and two protein-bound pools of diadinoxanthin cycle pigments in diatoms. Plant Physiol 154: 1906-1920.

Letunic I., Bork P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res 44(Web Server Issue): W242-W245.

Lohr M., Wilhelm C. (1999) Algae displaying the diadinoxanthin cycle also possess the violaxanthin cycle. PNAS 96(15): 8784-8789.

Lohr, M., Wihelm C. (2001) Xanthophyll synthesis in diatoms: quantification of putative intermediates and comparison of pigment conversion kinetics with rate constants derived from a model. Planta 212(3): 382-391.

Marchler-Bauer A., Bo Y., Han L., He J., Lanczycki C. J., Lu S., Chitsaz F., Derbyshire M. K., Geer R. C., Gonzales N. R., Gwadz M., Hurwitz D. I., Lu F., Marchler G., Song J. S., Thanki N., Wang Z., Yamashita R. A., Zhang D., Zheng C., Geer L. Y., Bryant S. H. (2017) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res 45(Database Issue): D200-D203.

Martin J. F., Gudina E., Barredo J. L. (2008) Conversion of β-carotene into astaxanthin: two separate enzymes or a bifunctional hydroxylase-ketolase protein? Microb Cell Fact 7:3.

Melendez-Martinez A. J., Mapelli-Brahm P., Benitez-Gonzalez A., Stinco C. M. (2015) A comprehensive review on the colorless carotenoids phytoene and phytofluene. Arch Biochem Biophys 572: 188-200.

Moskalenko A. A., Karapetyan N. V. (1996) Structural role of carotenoids in photosynthetic membranes. Z Naturforsch 51c: 763-771.

Peng J., Yuan J., Wu C., Wang J. (2011) Fucoxanthin, a marine carotenoid present in brown seaweeds and diatoms: metabolism and bioactivities relevant to human health. Mar Drugs 9: 1806-1828.

Petersen T. N., Brunak S., von Heijne G., Nielsen H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8: 785-786.

Robinson J. T., Thorvaldsdottir H., Winckler W., Guttman M., Lander E. S., Getz G., Mesirov J. P. (2011) Integrative Genomics Viewer. Nature Biotechnol 29: 24-26.

Santabarbara S., Casazza A. P., Ali K., Economou C. K., Wannathong T., Zito F., Redding K. E., Rappaport F., Purton S. (2013) The requirement for carotenoids in the assembly and function of photosynthetic complexes in *Chlamydomonas reinhardtii*. Plant Physiol 161: 535-546.

Shrestha R. P., Haerizadeh F., Hildebrand M. (2013) Molecular genetic manipulation of microalgae: principles and applications. In: Richmond A., Hu X. (Eds), Handbook of microalgal culture: applied phycology and biotechnology, 2nd edn. John Wiley & Sons, Ltd., Oxford UK, pp. 146-167.

Shrestha, R. P., Hildebrand M. (2015) Evidence for a regulatory role of diatom silicon transporters in cellular silicon responses. Eukaryot Cell 14(1): 29-40.

Sievers F., Wilm A., Dineen D., Gibson T. J., Karplus K., Li W., Lopez R., McWilliam H., Remmert M., Soding J., Thompson J. D., Higgins D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7: 539.

Smith S. R., Abbriano R. M., Hildebrand M. (2012) Comparative analysis of diatom genomes reveals substantial differences in the organization of carbon partitioning pathways. Algal Res 1(1): 2-16.

Smith S. R., Gle C., Abbriano R. M., Traller J. C., Davis A., Trentacoste E., Vernet M., Allen A. E., Hildebrand M. (2016) Transcript level coordination of carbon pathways during silicon starvation-induced lipid accumulation in the diatom *Thalassiosira pseudonana*. New Phytol 210(3): 890-904.

Thorvaldsdottir H., Robinson J. T., Mesirov J. P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14: 178-192.

Tian L., DellaPanna D. (2004) Progress in understanding the origin and functions of carotenoid hydroxylases in plants. Arch Biochem Biophys 430(1): 22-29.

Tran D., Haven J., Qiu W., Polle J. E. W. (2009) An update on carotenoid biosynthesis in algae: phylogenetic evidence for the existence of two classes of phytoene synthase. Planta 229(3): 723-729.

Wilhelm C., Buchel C., Fisahn J., Goss R., Jakob T., LaRoche J., Lavaud J., Lohr M., Riebesell U., Stehfest K., Valentin K., Kroth P. G. (2006) The regulation and nutrient assimilation in diatoms is significantly different from green algae. Protist 157(2): 91-124.

Wu H., Cockshutt A. M., McCarthy A., Campbell D. A. (2011) Distinctive photosystem II photoinactivation and protein dynamics in marine diatoms. Plant Physiol 156(4): 2184-2195.

APPENDIX 2.A BLAST RESULTS

1. Phytoene Synthase (PSY)

Previously reported single-copy PSY in *P. tricornutum* (Phatr2_56481) [Coesel et al. 2008], was confirmed to have no other BLAST hits in the *P. tricornutum* genome. A BLAST search against the *T. pseudonana* genome yielded two hits, one previously published as PSY1 (Thaps3_268908) [Coesel et al. 2008], and previously unpublished Thaps3_263269, which overlaps with previously published PSY2 (Thaps3_258309) [Coesel et al. 2008].

2. Phytoene Desaturase (PDS)

BLAST queries of the previously published *P. tricornutum* PDS1 (Phatr2_45735) and PDS2 (Phatr2_55102) [Coesel et al. 2008] against the *P. tricornutum* genome yielded each other, as well as a large chromosomal region (chr_1:926023-1979107) that includes Phatr2_53974, the ζ-carotene desaturase (ZDS) (3). In *T. pseudonana*, BLAST searches with the aforementioned *P. tricornutum* gene products yielded Thaps3_23291, which overlaps with the previously published PDS1 (Thaps3_6524) [Coesel et al. 2008], Thaps3_1383, previously published as PDS2 [Coesel et al. 2008], Thaps3_bd_1474 (not previously published), as well as the *T. pseudonana* ZDS (3), Thaps3_28432.

3. ζ-Carotene Desaturase (ZDS)

No candidates besides Phatr2_53974 and Thaps3_24832 (2) were found. The former had been reported by Coesel et. al. [2008], and the latter overlaps with the previously published gene model, Thaps3_37288 [Coesel et al. 2008].

4. Carotene Cis-Trans Isomerase (Prolycopene Isomerase) (CRTISO)

Exhaustive reciprocal BLAST searches between the *P. tricornutum* and *T. pseudonana* genomes starting with the products of four genes identified by Dambeck et al. [2012] as *P. tricornutum* CRTISO candidates (Phatr2_45243, Phatr2_9210, Phatr2_54842, Phatr2_51868) as queries yielded numerous hits in both organisms.

For *P. tricornutum*, the findings (in addition to the aforementioned genes) were Phatr2_54826, Phatr2_54800, Phatr2_42980. Three large chromosomal regions were repeatedly found as well: chr_1:1011072-1894441 (containing Phatr2_42890, Phatr2_9210, and Phatr2_53974, the ZDS), chr_6:586147-620230 (containing Phatr2_45243), and chr_15:52824-647530 (containing Phatr2_54826 and Phatr2_54800).

The findings for *T. pseudonana* were Thaps3_7094, Thaps3_21900, Thaps3_21847, Thaps3_5221, Thaps3_10233, Thaps3_11636, Thaps3_5859, Thaps3_25361, and Thaps3_10254. Two large chromosomal pieces were also found: chr_6:68923-1558412 (containing Thaps3_5859) and chr_15:676997-739441 (containing Thaps3_10233 and Thaps3_10254).

5. Lycopene β-cyclase (LCYB)

Phatr2_56484 [Coesel et al. 2008] generated no additional BLAST hits in the *P. tricornutum* genome, and only Thaps3_270357 in the *T. pseudonana* genome. The latter generated no additional BLAST hits in either of the genomes, and overlapped with the previously reported gene model, Thaps3_261407 [Coesel et al. 2008].

6. β-Carotene Hydroxylase (BCH)

As reported in Coesel et al. [2008], no BCH was found in the *P. tricornutum* genome, and only Thaps3_263437, previously reported as a partial sequence [Coesel et al. 2008], was found in the *T. pseudonana* genome.

7. LUT-Like (Lutein Deficient-Like) (LTL)

Exhaustive reciprocal BLAST searches between the *P. tricornutum* and *T. pseudonana* genomes, starting with Phatr2_50101 and Phatr2_26422 reported as LTL1 and LTL2, respectively, by Coesel et al. [2008], yielded many hits in both genomes.

For *P. tricornutum*, the findings were Phatr2_34027, Phatr2_33568, Phatr2_6940, Phatr2_46438, Phatr2_31339, Phatr2_47234, Phatr2_37006, Phatr2_43466, Phatr2_43467, Phatr2_43562, Phatr2_50619, Phatr2_43469, Phatr2_32833, Phatr2_43537. Several regions without available gene models also appeared in the BLAST results: chr_2:488287-489217, chr_4:1314546-1315166, chr_8:134705-134894, chr_8:983359-983442, chr_15:149534-149668, and chr_15:494943-495059. Where open reading frames were readily apparent, the hypothetical protein products were included as BLAST search queries. Additionally, there were several larger regions: chr_2:487571-543110, chr_2: 540661-973489, chr_4:48892-1315175 (contains Phatr2_34027), chr_8:134708-983593, and chr_11:518522-869181 (includes Phatr2_37006).

For *T. pseudonana*, Thaps3_9541 was confirmed to be LTL1, and a more complete model for LTL2 (Thaps3_270336) compared to the previously reported one that was missing the N-terminus (Thaps3_36235), was found [Coesel et al. 2008]. Additional results were Thaps3_33926, Thaps3_32491, Thaps3_1549, Thaps3_264647, Thaps3_25944, Thaps3_14875, Thaps3_4027,

Thaps3_4026, Thaps3_bd_518, Thaps3_269400, Thaps3_264325, Thaps3_263399. Two large chromosomal regions were also found: chr_9:88966-685832 (includes Thaps3_270336) and chr_3:2376984-2382496 (includes Thaps3_14875, Thaps3_4026, Thaps3_4027, and Thaps3_25944, which appear immediately adjacent to each other, in the order listed).

8. Zeaxanthin Epoxidase (ZEP)

For *P. tricornutum*, previously published [Coesel et al. 2008] ZEP1 (Phatr2_45845), ZEP2 (Phatr2_56488), and ZEP3 (Phatr2_56492) were confirmed. Additionally, Phatr2_43425, Phatr2_47925, Phatr2_45936, and chr_21:64775 – 64879 (no gene model) were found.

For *T. pseudonana*, Thaps3_270370 was identified as ZEP1, overlapping with the previously reported Thaps3 269147, and Thaps3_261390 was confirmed as ZEP2 [Coesel et al. 2008]. Additional findings were Thaps3_1961, Thaps3_6395, Thaps3_20663, Thaps3_22671, and chr_9:932102-933624 (no gene model).

9. Violaxanthin De-Epoxidase (VDE), VDE-Like (VDL), VDE-Related (VDR)

In *P. tricornutum*, previously published VDE (Phatr2_44635), VDL1 (Phatr2_46155), and VDL2 (Phatr2_45846) were confirmed [Coesel et al. 2008], and Phatr2_bd_1281 was found. Additional found sequences without available gene models were bd_29x34:989-1110, chr_2:9199-9410, and chr_1:974981-975118.

In *T. pseudonana*, previously reported VDE (Thaps3_7677) and VDL1 (Thaps3_22076) were confirmed [Coesel et al 2008], and Thaps3_11707 as well as chr8: 84698 – 842033 were found.

Only previously reported Phatr2_56450 and Thaps3_270211 [Coesel et al. 2008] were found in the VDR search.

APPENDIX 2.B ALIGNMENTS AND PERCENT IDENTITY MATRICES

1) CRTISO Percent Identity Matrix

1: Thaps3_25361	100.00	47.05	16.73	16.24	16.79	18.23	19.17	18.06	16.46	17.20	15.98	19.00	18.10	18.94	17.98	19.76
2: Phatr2_42980	47.05	100.00	16.61	16.85	16.60	16.04	16.96	18.22	16.96	17.09	19.25	18.32	19.62	19.64	17.65	19.75
3: Thaps3_5221	16.73	16.61	100.00	57.64	35.21	36.94	32.19	32.73	30.20	31.70	32.04	32.56	18.32	18.30	17.49	19.42
4: Phatr2_54826	16.24	16.85	57.64	100.00	36.55	35.94	34.01	32.73	31.65	33.06	32.34	34.94	17.92	20.03	18.49	19.77
5: Thaps3_21847	16.79	16.60	35.21	36.55	100.00	58.92	36.35	37.59	32.29	33.98	32.16	35.21	17.42	18.71	18.18	22.05
6: Phatr2_51868	18.23	16.04	36.94	35.94	58.92	100.00	35.51	34.78	31.49	33.59	31.83	34.95	20.00	19.75	18.37	20.99
7: Thaps3_7094	19.17	16.96	32.19	34.01	36.35	35.51	100.00	59.07	34.94	38.37	37.07	38.36	22.20	22.60	20.13	22.95
8: Phatr2_54842	18.06	18.22	32.73	32.73	37.59	34.78	59.07	100.00	33.74	37.45	35.55	38.13	21.05	20.56	20.48	21.98
9: Thaps3_10233	16.46	16.96	30.20	31.65	32.29	31.49	34.94	33.74	100.00	85.09	36.67	36.57	18.40	20.49	17.93	20.19
10: Phatr2_9210	17.20	17.09	31.70	33.06	33.98	33.59	38.37	37.45	85.09	100.00	37.65	38.39	19.05	21.46	20.10	22.53
11: Thaps3_21900	15.98	19.25	32.04	32.34	32.16	31.83	37.07	35.55	36.67	37.65	100.00	43.01	18.59	19.59	19.51	19.76
12: Phatr2_45243	19.00	18.32	32.56	34.94	35.21	34.95	38.36	38.13	36.57	38.39	43.01	100.00	20.35	22.35	20.78	21.30
13: Thaps3_11636	18.10	19.62	18.32	17.92	17.42	20.00	22.20	21.05	18.40	19.05	18.59	20.35	100.00	59.69	22.22	27.49
14: Phatr2_54800	18.94	19.64	18.30	20.03	18.71	19.75	22.60	20.56	20.49	21.46	19.59	22.35	59.69	100.00	21.68	24.68
15: Thaps3_5859	17.98	17.65	17.49	18.49	18.18	18.37	20.13	20.48	17.93	20.10	19.51	20.78	22.22	21.68	100.00	28.43
16: Thaps3_10254	19.76	19.75	19.42	19.77	22.05	20.99	22.95	21.98	20.19	22.53	19.76	21.30	27.49	24.68	28.43	100.00

CRTISO Alignment

```
Thaps3_25361      MRMGRPNKKLRSTSKQTTNPNPPKYSSPTLVVGQVSSNVIHSIYGPALTKLAVESV----        56
Phatr2_42980      MRIGKPLRKSRAWKKV-GVNGTPKYVSPESVVGRINSDIISSVLAPKIAALASASL----        55
Thaps3_5221       ------------------------------------------------------------        0
Phatr2_54826      ------------------------------------------------------------        0
Thaps3_21847      ------------------------------------------------------------        0
Phatr2_51868      ------------------------------------------------------------        0
Thaps3_7094       ------------------------------------------------------------        0
Phatr2_54842      ------------------------------------------------------------        0
Thaps3_10233      ------------------------------------------------------------        0
Phatr2_9210       ------------------------------------------------------------        0
Thaps3_21900      ------------------------------------------------------------        0
Phatr2_45243      ------------------------------------------------------------        0
Thaps3_11636      ------------------------------------------------------------        0
Phatr2_54800      --------------------MTPVTQSPPVELSTD-PPVALSLALPPLSPTADGTLQHHH        39
Thaps3_5859       ------------------------------------------------------------        0
Thaps3_10254      ------------------------------------------------------------        0


Thaps3_25361      ----------------------EEYADAV-LRWEASLPEVLVK-----------------        76
Phatr2_42980      ----------------------ERYAGEL-LVYEDIMKKVNSN-----------------        75
Thaps3_5221       ------------------------------------------------------------        0
Phatr2_54826      ------------------------------------------------------------        0
Thaps3_21847      ------------------------------------------------------------        0
Phatr2_51868      ------------------------------------------------------------        0
Thaps3_7094       ------------------------------------------------------------        0
Phatr2_54842      ------------------------------------------------------------        0
Thaps3_10233      ------------------------------------------------------------        0
Phatr2_9210       ------------------------------------------------------------        0
Thaps3_21900      ------------------------------------------------------------        0
Phatr2_45243      ------------------------------------------------------------        0
Thaps3_11636      ------------------------------------------------------------        0
Phatr2_54800      LTTTDDESSSPWPVIRSVFRGQNNFSDPLNRGWNPWRPGISSRQDKCGVEYVKMHGQYFP        99
Thaps3_5859       ------------------------------------------------------------        0
Thaps3_10254      ------------------------------------------------------------        0


Thaps3_25361      ---------P------SQLDDAEDIDVDADGTFEKGEEVEVDLDGSILPSHDNDDDKTSS        121
Phatr2_42980      ---------E------S------------IDGLLESDS--------SIIIQ----DG---I        94
Thaps3_5221       ------------------------------------------------------------        0
Phatr2_54826      ------------------------------------------------------------        0
Thaps3_21847      ------------------------------------------------------------        0
Phatr2_51868      ------------------------------------------------------------        0
Thaps3_7094       ------------------------------------------------------------        0
Phatr2_54842      ------------------------------------------------------------        0
Thaps3_10233      ------------------------------------------------------------        0
Phatr2_9210       ------------------------------------------------------------        0
```

```
Thaps3_21900      ------------------------------------------------------------        0
Phatr2_45243      ------------------------------------------------------------        0
Thaps3_11636      ------------------------------------------------------------        0
Phatr2_54800      TSGSGFSTGPVGQHDGHELEQRCDTERQGGGHVNKGNHVEANVATSVLCATE-----CCL        154
Thaps3_5859       ------------------------------------------------------------        0
Thaps3_10254      ----------------------------MGFIVRGTRARSVVS------SR-----LAV        20

Thaps3_25361      PTMPTSNRL--FTTQSSIDNLTALLTDTSQH-FSTTNAWKIHANAAKFERLLDEKYGRFR        178
Phatr2_42980      PQQ-PQRPV--FHSQFSVDEAASIFSETSEHFFAKAGKWKAHANAAKFERILDEKYGILR        151
Thaps3_5221       ------------------------------------------------------------        0
Phatr2_54826      -----------MLV-------------------ESKKSRDG----SRTSS----RS----        18
Thaps3_21847      ------------------------------------------------------------        0
Phatr2_51868      ------------------------------------------------------------        0
Thaps3_7094       ------------------------------------------------------------        0
Phatr2_54842      ------------------------------------------------------------        0
Thaps3_10233      ------------------------------------------------------------        0
Phatr2_9210       ------------------------------------------------------------        0
Thaps3_21900      ------------------------------------------------------------        0
Phatr2_45243      ------------------------------------------------------------        0
Thaps3_11636      ------------------------------------------------------------        0
Phatr2_54800      PHREPEPKTVCVVVLRSLQRDPTFQQDEVSAFLGAFRGWIGNEVDRLFEW----AA----        206
Thaps3_5859       ------------------------------------------------------------        0
Thaps3_10254      KGQSPNPL-----------------------------------ESQLTNW----RE----        37

Thaps3_25361      PFIESHPELEVFIKKVQRKYAMGQFSPLRKGEGPMSTTSSIMLLFMMHRNGVRKELVALV        238
Phatr2_42980      PFITNHPEIEHFIRGVQRKYAMGYFSPFRQGDPPIPRSTAVIILFMMQRGQMRWEIMLLT        211
Thaps3_5221       ------------------------------------------------------------        0
Phatr2_54826      --------------LDTKTHCVCSTS------------KQ----------------------        32
Thaps3_21847      ------------------------------------------------------------        0
Phatr2_51868      ------------------------------------------------------------        0
Thaps3_7094       ------------------------------------------------------------        0
Phatr2_54842      ------------------------------------------------------------        0
Thaps3_10233      ------------------------------------------------------------        0
Phatr2_9210       ------------------------------------------------------------        0
Thaps3_21900      ------------------------------------------------------------        0
Phatr2_45243      ------------------------------------------------------------        0
Thaps3_11636      ------------------------------------------------------------        0
Phatr2_54800      --------------TAAKALCMANWQ------------GKSRLLK--K---FGATVRRLD        235
Thaps3_5859       ------------------------------------------------------------M       1
Thaps3_10254      --------------PPRRPFSKVQTH------------ASNQL------------QSEVM        59

Thaps3_25361      ALFTLVGLEPWALVGLVCVGKYSVDQRRRKRIGGMP--------------KKVKV----V        280
Phatr2_42980      TLFFLIGLQPWALVAVVGVLQGLLMRRKAKPLGKMK--------------RFIPA----V        253
Thaps3_5221       ------------------------------------MKLLF-VASTLIGVLSFTPP--QVL        22
Phatr2_54826      ----------NSVRPANTLLRSR-------ARHTLIWLVFYVWEWNTTTTAFAPSPSRIA        75
Thaps3_21847      ----------------------MK-------VSTTAT-FALLQIGTAAVSAFTSP-----        25
Phatr2_51868      ------------------MANISKD-R--L-TGRFLA-FLLLVLANKETSSFCVQSGYRS        37
Thaps3_7094       ------------------MLPHT-------TVHGVVALATLLLNAFVLVDSFAPS-----        30
Phatr2_54842      --------------------MFAISSQLT--LTLVGHLILLHMM-ENSAICSAFVPASQRTT        39
Thaps3_10233      ------------------MIGRKY--S--LAASA----LAIIA-SLTSTTAFAPPSSSLL        33
Phatr2_9210       ------------------------------------------------------------        0
Thaps3_21900      ------------------MIR---------SISA----LALLAACCPSVFSFAPLSVF--        27
Phatr2_45243      --------------------MR---------FSER----SLIACAICSISSAFVPIIHTPQ        28
Thaps3_11636      -----MNT------ITDLLFQNPSTFIV-L--------LPLLFIASFIFYIT--------        32
Phatr2_54800      KACGPFGWTAWILFPCQTLMQRRDHYTR-YAIGRS-R-TLGDAESCFLVWLF--------        284
Thaps3_5859       ELFSSINYDPWTLVPAS--------FYY-PTIITA-C-IPLLFIATAYWLLIRRA-----        45
Thaps3_10254      DSLSKISS---SLLGDGKHT---KVTTV-ATVAGL-T-LGTLFIARRIYLS---------        101

Thaps3_25361      ESYYAHGVVGE----------EEEESEEVERSK----KYAILEKPVGT-------IFNPA        319
Phatr2_42980      ESYYTDAKTDT----------EK---------------HELLLHPVGE-------PL-PS        280
Thaps3_5221       T--D-------------------------TRHR-----------P----PALCNSGDATN        40
Phatr2_54826      AFRA-------------------------SRGR-----------KLTTSVSSLVSGDKRD        99
Thaps3_21847      ----SI-------------NSVIRSPSTHLRS-----------SPSATAST---------        48
Phatr2_51868      RHYFSA-------------NFLSVQPSDVARG-----------SSTAPAAAI---ADAPT        70
Thaps3_7094       -PRCSH-------------------------R------YH----I---SSAA---STTLH        48
Phatr2_54842      FSNCRR-------------------------S------KN----RVGRHGCF---LLASQ        61
Thaps3_10233      RSTRLHSTVEETTNGEAATNTNVEQIKDTSRDKVMTFSYDMSIEPKYEKPTY---PGTGN        90
```

```
Phatr2_9210      ------------------------------------------------------------        0
Thaps3_21900     ---------------------------------------------RANAP-------SSL        35
Phatr2_45243     -----H------------------QSPRTTRH---QFT------RIYAAV-------SSV        49
Thaps3_11636     -------------------RWPQARP-VQFRR-----------------AD--RFRPEKV        53
Phatr2_54800     --------------------HWPARRVKLHPRR-----------------AS--RFRPELV       306
Thaps3_5859      ---QLH-------------REKGKLPKYDAIP-----------------SSVLKHIASKQ        72
Thaps3_10254     --------------------WMKEFPSSDSLP-----------------STNPVKQ-GFS       123


Thaps3_25361     DLSLRDEEYDVILLGCGPEVLYTASLL-SRAGKKTLVLSPREDASGCLTLQNG-------       371
Phatr2_42980     KEEIDASLFDALILGSGPASLYIASLL-SRAGRKVLVLSSRNDASGCLSIKHAE------       333
Thaps3_5221      DGGDEVHEVDIAVVGAGIGGLCAGAILNTLYDKKVGVYESHYLAGGCAHSFSRSVK----        96
Phatr2_54826     CASPADDLVDVAIIGAGLGGLCAGAILNTLYGKKVGIYEAHYLAGGCAHAFDRRAA----       155
Thaps3_21847     ITDADEEEWDVVVVGSGVGGLSAAAMC-ARYGLKTICVEAHDAPGGVAHSFERRAS----       103
Phatr2_51868     GSIYREETVDVVVIGAGVGGLSCAALS-SKYGMDTLCLEAHDTAGGCAHSFERYSA----       125
Thaps3_7094      ATTTPHSEYDAIIVGSGIGGLSAAALL-SHYGYSVAVFEAHSTPGGAAHGYTVNA-----       102
Phatr2_54842     SATPGTTSPTPVVVGSGIGGLCAAAML-AKYGYTVAVLESHNVPGGAAHGFTARDP----       116
Thaps3_10233     GMSGDSGEYDIIVIGSGMGGLACSALS-AKYGSRVLCLESHIKVGGSAHTFSRMHN----       145
Phatr2_9210      ---------DVIVIGSGMGGLACGALS-AKYGDKVLVLESHIKCGGSAHTFSRMHN----        46
Thaps3_21900     ASTTYQDEVDCIVIGSGIGGLSCAALL-AATGRTVRVLEQHYEIGGCAHAFYMDMNGKTV        94
Phatr2_45243     PSNSIPDEADVVVIGSGLAGLSCAALL-AHCGKRVVVLESHDAPGGAAHGWE--------       100
Thaps3_11636     P-----SNIDTIVIGSGSGGSTVANLL-AQSGQRVLVLEQHSVTGGCTHSFR--------        99
Phatr2_54800     LENGKQRRFDTIVIGSGSGGCACANLL-AQSGQRVLILEQHTKTGGCTHSFR--------       357
Thaps3_5859      VLRDLSGKIDVAIVGSGIAALSNASAL-AHQGYKIAVFEQNEIVGGCTHTFE--------       123
Thaps3_10254     IKSVSSTNWDVIVIGSGAGGLTTAALL-SKEGKKVLVLEQHDIAGGNLHTFS--------       174
                             ::*.*      .    :  .      . .   .*


Thaps3_25361     ------------KTNVPFDIDGSNIAHLARQ---------QSLLAPA------LCTTTDT       404
Phatr2_42980     ------------YSNVPFDVEASNVAKISRQ---------QQILAPA------LCTETDT       366
Thaps3_5221      ------IGDDEQPTTFTFDSGPTIVLGCSK----EPYNPLQQVLRAVGVDDQIEWLPYDG       146
Phatr2_54826     ------D-----GVNFTFDSGPTILLGCSS----PPFNALQQVLDAVGQK----------       190
Thaps3_21847     ---------SSPNRPFVFDSGPSLLSGMSS----KGTNPLRQVLDAVGTADDIDWVTYDG       150
Phatr2_51868     ---------ASKTTPFRFDSGPSLVSGLSE----KGTNPLRQVLDAVGTAEEVQWKTYDG       172
Thaps3_7094      ----------KDVGPLTFDTGPSFFSGLNSNYPAKSSNPLRSILDIID--EKVECIPYTT       150
Phatr2_54842     ----------KIEGEFRFDTGPSFFSGINSDTPAKASNPLRTVLDAID--ERVECVPYTT       164
Thaps3_10233     ------------GGKYSFEVGPSIFEGLDR----PSLNPLRMIFDILE--ETMPVKTYKG       187
Phatr2_9210      ------------GEKYSFEVGPSIFEGLDR----PSLNPLRMIFDVLE--EEMPVKTYTG        88
Thaps3_21900     PSSALKDDPTKKGELFHFEAGPSLYSGLSEE---RTPNPLKHIYQMIE--EEPEWLTYDQ       149
Phatr2_45243     ------------RRGFHFESGPSLYSGFAME---RSPNPLKNIFQITG--EDCEWITYDR       143
Thaps3_11636     ------------EEGCEWDTGLHYVSKAMA----TPTKRAGAIMSFMS-RGKQSFTPFPT       142
Phatr2_54800     ------------DRGCEWDTGLHYTSAGMG----RSTCRPGAIMHFMT-QGLQKWTPL--       398
Thaps3_5859      ------------KQGFEFDVGVHYVGGFG--------TVVKHMYDELS-DGQLKWTKL--       160
Thaps3_10254     ------------EKGYEFDTGLHYVGGKVG----DKSSSVRKQLDYVM-DTDVEWEKM--       215
                                 ::


Thaps3_25361     QGGIRFA--------RIG-SEVDGYAHSILSVPGLGTDSISNECIPIVLT-----AEGEV       450
Phatr2_42980     QGGVRFA--------QIG-SNEDAHAFEILSIPGMGTDSYDEELPFILNA-----DGGTA       412
Thaps3_5221      WGMIEHPMQ----------PKEKRWKF--KV----G----PNHFEDGPLQVF---ASNLN       183
Phatr2_54826     -----NPGK----------DNELRWKV--IL----G----RDEFQRGPLTRF---GG-PK       221
Thaps3_21847     WMVHDTAFP----------MDDSRSSFRLTT----G---SDGTWEDAIEAKAG--VDSRR       191
Phatr2_51868     WLVHDTS---------------DDKVFKVTT----G---DSGAFEDALEKKAG--INAKR       208
Thaps3_7094      FGLMFPE----------------GVFVHSSNF----G----KEG--STVEA-----VSGSN       180
Phatr2_54842     FGLQFPE----------------GNFEHSCFF----G----AQG--GLLEQ-----LQGTT       194
Thaps3_10233     LGYWTPS----------------GYWRFPIGS---------REGFEQLLMEQCG--EDGEK       221
Phatr2_9210      LGYWTPT----------------GYWRFPIGS---------QSKFEDLLMEQA---EDGPK       121
Thaps3_21900     WGAFLPE----------------APEGYQMSI----G----AENFCKILET-----YGGEG       181
Phatr2_45243     WGTVMPD----------------GT-KFAAKI----G----PEEFQDVLES-----QGGPG       174
Thaps3_11636     STPYDEIVFPKDANVKDGAPNEFSHKF--YD----G---VNRTVSSVIGSIDPSDNELKH       193
Phatr2_54800     QDPYDEVIFPPDDFVKLGVPNESSYRF--VS----G---ADETIQSVLASIDPEHRELEK       449
Thaps3_5859      DRVYDVMYNGRTG----------ERYEI--TD----D---HDK-----------NRRVLTK       191
Thaps3_10254     DDIYDVAICDEEQ----------F--NF--CS----S---WKT-----------LKVELKK       244


Thaps3_25361     ALAEYCSTYLGDAFPGTDLDGNDNGNSTSLSYLKACGQINAGSGDFYL-------AKLFP       503
Phatr2_42980     GLIDDAAKYLNDGWPDAE---GGNGNSVTGAYAAACEAINSTANEFYI-------SKILS       462
Thaps3_5221      ALEEF------------------------NQLREITKPLVTGAATIPAMAMRPGQSALV-       218
Phatr2_54826     ALEEF------------------------EALREATKDLLAG-AKIPAMAMRPGPSALV-       255
Thaps3_21847     EFTKF------------------------KKKMMSSGGLSESSALLPPMALRGDFGALF-       226
Phatr2_51868     EFIEF------------------------KRKVLEEGGLAEASAYIPPFALRGGITALA-       243
Thaps3_7094      GVQEW------------------------ASLMKSMDPLAQAVDAMPTLALRADLGLLA-       215
Phatr2_54842     AQKEW------------------------QALMQSMGPLEKAVAALPTAALRGDIGLLL-       229
```

```
Thaps3_10233      AIGEW------------------------KALRERLRTLGGSTQAVALLNLRQDAGFLA-        256
Phatr2_9210       AVEEW------------------------NMLRKRLKTLGGSTTAVSLLNLRQDPGFLA-        156
Thaps3_21900      AVEDW------------------------EKLAEQLRPMAGGIKGIPHAAIRGDWGIFL-        216
Phatr2_45243      AREEF------------------------AALMERMKPLSDAAQALTSLALREDPAVVV-        209
Thaps3_11636      RVDTF------------------------MDI---CLDVHNG----F-VAL--GIYRLLP        219
Phatr2_54800      RARLY------------------------MDL---CTDINSG----F-TAL--GISRVLP        475
Thaps3_5859       DFGID------------------------EQSW-RKFDRKKAYAKFWAMVV--FSLKLFH        224
Thaps3_10254      KFPEE------------------------SDAIDKHFQLVQSTVKLFPVFM--GIKNLPT        278
                                                                          .

Thaps3_25361      KAAES--------FKSSDSNV-----YQQ-ASIRPASTFL-----N--KCLPLNTHVRAL        542
Phatr2_42980      EKVNS--------LRS--SPT-----YQD-SGIRYAQSFL-----N--KTFTINPHTRSL        499
Thaps3_5221       PLLRY---------LPSLISI---ISNGVEASTGPFAPYM-----NGPIFTVKDPWLRSW        261
Phatr2_54826      PLIRY---------FSTLVTL---LSQGSK-ATGTFASFI-----DGPNFTVTDPWLRSW        297
Thaps3_21847      T-MGS---------YVFKFLT---IGLQGTLLTGPFTECM-----N--LYGLNDRFNRQW        266
Phatr2_51868      S-LAN---------YMFKLLS---IGSKGALLTGPFSKVM-----D--LHGLKDPFVRKW        283
Thaps3_7094       STSQF---------LPNFAKL---NPLQNLKLTKPFSNII-----N--EAGVKDTFIRNW        256
Phatr2_54842      TAAPF---------LPNFTTL---NPLENLKLTQPFSAIV-----N--P-SVSNVFTRNW        269
Thaps3_10233      TTAGS---------LPFVVTH---PDVFG-DLSLTFDDLS-----KTVDEFVTVPFLRNF        298
Phatr2_9210       TTAGS---------LPFVATH---PDVFL-DLSLTFDSLH-----KTVDKIVTVPFLRNF        198
Thaps3_21900      TLILK---------YPLSFMN---VLKYAPAFTAPFD--L-----D--KLGVTNKFLRNY        255
Phatr2_45243      -TLLK---------YPRDLIA---TLAQGQALNEPFKNIM-----D--EMKIENKFVKNW        249
Thaps3_11636      SYLKFLMKDKVERLYKYGSMTVKDAQHAVLKLGYSKEELLK-NCPTAP-EMEDDPSIRRM        277
Phatr2_54800      SWMHFLVRSRIDRLMKFAAMTVRDVQYGMLNLGLTIEELLKDGCPPAPAGSEPDPSIRRL        535
Thaps3_5859       PMVLRLA-------WPFVC-----IPYRRCALRSTIDVLI-----------NDCGFSQEA        261
Thaps3_10254      PLFRLVM-------WLFDS-----K---LGVYRKTTKEVL-----------ESITSNRKL        312
                                                                          :

Thaps3_25361      MAAI--GMANENLSPDKTSMAAHVTNVCAMTSTEGYA--------YPVGGPRALCHALTS        592
Phatr2_42980      MAGI--GMKGENIRPGATSMAAHVTNISAALSGEGMH--------YPIGGPRALCRALAN        549
Thaps3_5221       LNALAFSLSG---LPADRTSAGAMAYVLFDMHREGAA------LDYPRGGLGEVVKALVN        312
Phatr2_54826      LDALAFSLSG---LPASRTAAAAMAFTLSDMHRPGAA------LDYPKGGMGAIAEALVR        348
Thaps3_21847      FDYLAFALSG---LDAAHTQAAPVAYTMIDLHKDGAV------LDYPKGGMDSMIQALVN        317
Phatr2_51868      FDYLAFALSG---VDASHTQAAAVAYMMMDLHKKDAV------LDYPMGGMDSLVQALVS        334
Thaps3_7094       LDVLCFCLSG---VPSDGTITAEMAMMMGEFYDEDAI------MDCPVGGASAIVDALVR        307
Phatr2_54842      LDLLCFCLSG---LPAKGTITAEMAMMMGEFYAPGAV------MDCPKGGAQSIVKALVR        320
Thaps3_10233      IDTMCI-FCG---FPAKGAMTAHLLYILERFFEETAA------FSVPIGGTCELGNTLQR        348
Phatr2_9210       IDTMCI-FCG---FPAKGAMTAHMLYILERFFEESAC------YSVPIGGTCEMGNTLVR        248
Thaps3_21900      LEMLAFLLQG---LPADQTLTVVMAYMVEDFFRENAV------MDFPKGGSGELMGALAR        306
Phatr2_45243      LDMLCFLLQG---LPASDTMNAVMAYMLADWYRPGVT------LDFPKGGSSSIVSALVR        300
Thaps3_11636      TAVLTHPIGDYAVQPRDATF--AAHGVTMAHYVNGSPNHNLVITKHTVGATQNISTRLTS        335
Phatr2_54800      KAVLTHPIGDYAVQPRDATM--AAHGVTMAHYQDGAC--------YCVGPTQQISVRSSS        585
Thaps3_5859       AGALTYHWGDHVVPPHRCPF--FMTALLDTHYKGGGY--------FPRGGSRSIAKCLVS        311
Thaps3_10254      QGVLSYHYGDYGEHPSRGAF--VMHSMICVHYRGGAY--------YPVGGPLSIAKSIAT        362
                     :     .                                   *    :

Thaps3_25361      VIEQ-----NGGRVVSGVLLQELLFEKLEKKEPKEETKDGES-KEPKPRCKGIRLENGL-        645
Phatr2_42980      VVLR-----SGGRVLTSVDVAELIFGEPREQASKGKQKEGDNDGPPPPRCVGVKLSDGR-        603
Thaps3_5221       GVEQK---SIGSKVHLSRHVESIDTNEE---------------G-DR--VIGLTVRKNGG        351
Phatr2_54826      GVQQG---SNGSQVHLRQPVEKIDFSED---------------G-TI--ATGLTLRNGR-        386
Thaps3_21847      GLEMKRDNVESGELRLKSRVERFVLNEV---------------K-NKATCTGVVLEN-G-        359
Phatr2_51868      GIKT-----NGGELRLNSRVERMILEDN---------------N-GRVECKGVVLTD-G-        371
Thaps3_7094       GIEK-----KGGKVFCNSRIDEICIENG---------------K-----AVGVRLAKNY-        341
Phatr2_54842      GIEK-----YGGEVVCNTHVQEIVVENE---------------K-----AVGVVIKQGK-        354
Thaps3_10233      GLEK-----YGGKLQLNAHVDEILVENG---------------R-----AVGVRLMN-G-        381
Phatr2_9210       GLEK-----FGGKIQLNAHVDEILVENG---------------R-----AVGVRLKN-G-        281
Thaps3_21900      GVTKR----EGCSVEVSTSVDEVIVENG---------------R-----AVGVKLAKSG-        341
Phatr2_45243      AVQK-----NGSSVCVNSHVDEILVENG---------------K-----TVGVRLTD-G-        333
Thaps3_11636      MVRS-----FGGEALIDATVRGIIIENG--------------------RAVGVKVSNTD-        369
Phatr2_54800      MVRE-----FGGEVLTDATVREIILEHG--------------------RAVGVRVSNTS-        619
Thaps3_5859       AITR-----RGGHVFALSPVDEILTKKN---------------MFGKFIATGVSVRGID-        350
Thaps3_10254      TIEK-----HGGKVLVRAPVSSVLVDEK---------------N----RAYGVVVKGKE-        397
                   :        .       :  .   .                      *: :

Thaps3_25361      --ELSVS-D-------------KGAVVSFMGMIPTFLQLVSPDVRTAEGV------PAGL        683
Phatr2_42980      --EIKFA-SDRFDE---KNGSCLPAVISMEGFIWTFINMLPDDIRMKYKV------PRGL        651
Thaps3_5221       KKVIVKA---------------KEGVVCNVPMWSLRKLIKNRNALSVLGGDKATSSSSGL        396
Phatr2_54826      ---RILA---------------REGVICNAPVWSLKSLRPTR------------------        410
Thaps3_21847      --TILKA---------------RRGVICNAPLWNMAKLLEDSITNPLDL-----------        391
Phatr2_51868      --TVVNA---------------RKGVVSNAPIWNMARILEDSVPGEVND-----------        403
Thaps3_7094       --SRIKA---------------TKGVISNLSVWDLMNSGIV---D---------------        366
```

```
Phatr2_54842      --QRVAA---------------SKAVISNLSVWDLFGSGIL---D---------------        379
Thaps3_10233      --NVVKA---------------RKAVVSNATPFDTVKLMPKAEGEPKG------------        412
Phatr2_9210       --NVVKA---------------NKAVVSNATPFDTVKMLGEKQALPEG------------        312
Thaps3_21900      --RIIKA---------------KEAVISNADLYNTYKFVPEGKHEGFD-KER---------        375
Phatr2_45243      --RKVHA---------------TQAVVSNADPYISNKLLLNARKSGQLNKAA---------        368
Thaps3_11636      --ELEECTSEEDLAKVPAV---------------EYNKFLPQDLPVV-------------        399
Phatr2_54800      --ALAECKSDAERAQVPVTELRAKAVVCATSVYNLYNNLLPQDLAQV--------------        664
Thaps3_5859       --IVVKKC-----------------VVSDAGFLNTFGIDSEGNKPALVDSNAAASQR---        388
Thaps3_10254      --VLAK-T-----------------IVSSIGAPATFGKLLPESHRHLV------------        425


Thaps3_25361      PALEERRPLMRVMISLKGNKDDLNLTGADWYRLPNATLPRDELDPMTGQVKFGTIGVDDD        743
Phatr2_42980      PALSSRRPVFKVLFALKGSADQLNVTGADYYRLPNAAVARDEFDQSSGQIKHGEIGWSDS        711
Thaps3_5221       KAKQSWMTSFDTD-PSTGRGSVLRPKPAEDTTIEKSLLEKCDSAEMTGSFLHLHLALNAT        455
Phatr2_54826      ----------------------------SGEADETLLGACDTAEKTGSFLHLHLALESS        441
Thaps3_21847      ----------------------------SVAAAVNDVRSQANEMEMTGSFMHLHLGIPND        423
Phatr2_51868      ----------------------------ARRSIVKAIQKQADDMSMTGSFMHLHLGIPKA        435
Thaps3_7094       ---------------------------TDLF--PEDFVKERKATPACPSFMHLHVGFQIT        397
Phatr2_54842      ---------------------------TTLL--PNSLVQQKLSTPLGKSFMHLHVGFRMS        410
Thaps3_10233      ---------------------------------LTKWREELGKLPRHGAISHLFLAIDAE        439
Phatr2_9210       ---------------------------------VAKWKEELGKLPRHGAIMHLFLAIDAK        339
Thaps3_21900      ---------------------------IEYL--GLTAKPKDGSVPFCKSFMHLHLAVKAE        406
Phatr2_45243      ---------------------------TDHLDALINTDKTEGGIADLKSFIHIHAGIDAA        401
Thaps3_11636      ----------------------------------KKFKDE--ATIRQSNGHVFLFCKLRGN        424
Phatr2_54800      ----------------------------------KEFQDPEKRTIQQSNGHIFLFCKIKGD        691
Thaps3_5859       ----------------------------ALLHNAKGFPTLDSVTPCISNLSLFIGLDRT        419
Thaps3_10254      ---------------------------------SKQLESMKDNMIASNLTLMSMFVGISDP        453


Thaps3_25361      NTGASEELILGEATDETEATT----SHTRGKRNKAATSKAPRSKFTSGVSWMKVSFPSAK        799
Phatr2_42980      DTGDNGDAYADGGKNLMDVINQDPGSISDEHIVNSSRKRARKTKFEAGSSWLHVSFPSAK        771
Thaps3_5221       GLDLQS---LE-PH-YTVMDRGL---EGDG-KV--IDG--VKDDSSGELNMIAVSNPCVL        502
Phatr2_54826      GLNLDN---LE-AH-YTVMDRSL---GGDG-SS--VNG--VLDGPCGILNMIAVSNPCKI        488
Thaps3_21847      GLPA----DLD-CH-HSVLNLEH---------------------DVTAAQNLVIVSIPTIF        457
Phatr2_51868      GLPE----HLE-CH-HSVLNMQD---------------------DVTAEQNMVIISIPTVF        469
Thaps3_7094       KEELSK---LQ-AH-YIFMNDWE---R----------------GVTAEENCALVSIPSVH        433
Phatr2_54842      KGELQT---LQ-AH-YMHMEDWG---R----------------GVQDEDNAVLVSIPSVH        446
Thaps3_10233      GLDLSHI--QD-PA-HLVVQDWD---R----------------SLQDSQNLCSFFIPSIL        476
Phatr2_9210       DLDLSHI--QD-PA-HLVVQDWG---R----------------SLQDSQNLCSFFIPSLL        376
Thaps3_21900      LIPE----DAP-PQ-WTVVQDWD---K----------------GIDATGNVVVVSVGSKL        441
Phatr2_45243      GLPDQPSADFP-AQ-WAVVRDWD---A--------------PEGVESPRNIVLCSMPSLI        442
Thaps3_11636      ADEIG----LP-DHNLWYFNGYD---LDDA-FDKYFAN-----PTEVRPPTVYIGFPCTK        470
Phatr2_54800      PTELK----LP-AHNLWYFNSYD---IDDA-FEAYFTD-----PVGQRPPTVYIGFPCTK        737
Thaps3_5859       DEELE----LP-AQNVWHVHDWD---HDAA-WKNMMNAISPYQSLADQTPFLFISNESAK        470
Thaps3_10254      ENSLA----LP-KRNYWIHDSWD---HDKN-IE-------AFKKNPTKPPVFFVSFSSAK        497


Thaps3_25361      ------DPSWQDRHGDVSTCVVTVEA-DDDFVQMFDTKPKIYSV--------------LK        838
Phatr2_42980      ------DPSFEERHGKTTTCVVTIEA-DDDFVTYFDTKPKIYVI--------------KN        810
Thaps3_5221       ------DNTLAPEGFIIMHAYG---AGNEPFEIWKPPTASKGNASPNTAGEGEIIGGERC        553
Phatr2_54826      ------DNSLAPDGTIVVHAYS---AGNEPYEIWEGLDR----------------------        518
Thaps3_21847      ------DPSLAPEGYHIIHAYT---AASEDFADWERMLIGELDGGK-------PEFTDYK        501
Phatr2_51868      ------DPSLAPEGYHVVHAYT---AACDGFDQWTPYLDSGKETG--------------K        506
Thaps3_7094       ------DNTLAPDNHAVLHIYT---PATELYERWENVKRNT----------------------        465
Phatr2_54842      ------DDTLAPEGYAVLHIYT---PATEDFTRWENVQSK-----------------------        477
Thaps3_10233      ------DKTLCPEGKHVIHVYS---SGGEPYEPWEKLTPGS-----------------------        508
Phatr2_9210       ------DKTLCPEGKHVIHVYS---SGGEPYEPWEKLKPGT-----------------------        408
Thaps3_21900      ------DQSLAPPGYHVIHAYT---AGNESYEDWEQFEHLMDDAA------------VRD        480
Phatr2_45243      ------DPSLAPEGKHVLHAYV---PATEPYADWAGMDRKS-----------------------        474
Thaps3_11636      FGLIRQDITWQKRFPNVSNCILISDGLYEWFEQWSDKPV----------------------RN        511
Phatr2_54800      ------DTSWKQRFPGVSNCILISDGLWEWFEKWQDKPV----------------------HN        772
Thaps3_5859       ------DPDFGTKHPGKATSEVFAVCKYDLFEKWADTAH----------------------NS        505
Thaps3_10254      ------DPTYSSRNPGKQVALVVGPGFFDHVAVFQNERV----------------------KH        532
                           *                      :


Thaps3_25361      A---GNGERERLRDRVLKDLLETFPQLQGQLE----TVQICGPVR---------------        876
Phatr2_42980      A-SATKGDLDRLLERVKKDVYHIFPQLRDKVD----HCEICGPFQ---------------        850
Thaps3_5221       SPSTYQALKDSRSKVLWRAVESVIPDARERTV----LALIGSPRT---HERFLRRPCG-S        605
Phatr2_54826      RSDGYMCLKEDRAEVLWRAVESIIPDARNRVV----ISEIGSPIT---HERFLNRPRG-T        570
Thaps3_21847      RTKAYKDLKQEKAEALWLALERIIPDVRERAKREGSVVEVGTPLT---HRRYNRRYRG-T        557
Phatr2_51868      VVDGYNELKDEKADVLWRAVERVIPDVRLRAKQKGSIILVGTPLT---HRRYNQRYRG-T        562
```

```
Thaps3_7094       --PEYNQLKEERSAFLWKVLEKIIPDIRQRAV----HSKVGTPLT---HQRFLNRYRG-S        515
Phatr2_54842      --EAYEKLKEERSQYLWKVLTRIVPDIRERAR----IVRVGTPLT---HQRFLRRYKG-S        527
Thaps3_10233      --EEYEAYKNERAEVLWRAVERCIPDVRDRVE----FSIVGSPLA---HEAFLRRDRG-T        558
Phatr2_9210       --QEYDDYKNERAKVLWEAVERCIPDVRDRLE----FSIVGSPLA---HEAFLRRDRG-T        458
Thaps3_21900      KDAAYQTFKDERAQPIWDAIQKRAPAVVKGA-C--VIEKVATPLT---HARFLNRHRG-N        533
Phatr2_45243      --EEYTKKKEQAADFLWSAIEEYIPNARDRAVP--GTVQIGTPLT---HERFLRRTRG-T        526
Thaps3_11636      RGEEYLEFKDKLTHHLLDILQEFVPQVKGRIE----YHHLGTPLS---EETFLASYRGGS        564
Phatr2_54800      RGSDYEEFKEKLSKHLLEILFEFVPEVKDKIE----FSFLGTPLS---EQTYLNSFCAGS        825
Thaps3_5859       RGDDYTELKEKIIESYLNVFYLHFPKTKGHEG----NLAKCTVLTMVSADSMA*------        554
Thaps3_10254      RGKEYTDMKKEWEIVYMEAFLKQFPELKDKVD----YVEFGTALS---NDFYLGTNRGAV        585
                            .         .    *

Thaps3_25361      SGLTHNGPRF-----------AIKGNRPETPYPGLYIGGADLTVGDSFSGAIVGGWLAAN        925
Phatr2_42980      KGLSHNPERF-----------AAKGIRADTPYPGLFVGGSDLTVGESFSGDIVGAWLAAN        899
Thaps3_5221       YGAAFED----------------CLKDGSTPISNLVLSGDGVF--PGIG-------IPAV        640
Phatr2_54826      YGSATED----------------YLADGSTSIGNLLLAGDGIF--PGIG-------LPAV        605
Thaps3_21847      YGPAPSNGND-----------VWELPGPKTPIEGLLACGDCCF--PGIG-------LPGV        597
Phatr2_51868      YGPAPGPGKD-----------VWELAGATTKIKGLLACGDSTF--PGIG-------LPGV        602
Thaps3_7094       YGPAIRAGDA-------------SFPFPNTPIQGLLLCGDSCF--PGIG-------VPAV        553
Phatr2_54842      YGPAIQAGVG-------------SFPFAGTPIRQLLTCGDSCF--PGIG-------VPAV        565
Thaps3_10233      YGMAWAAGSSAPQSGILGSVLPFPFPNLKTPVDGLLRCGDSCF--PGIG-------TPSA        609
Phatr2_9210       YGMAWAAGTSAPQAGLLQNILPFPFPNLKTPVDGLLRCGDSCF--PGIG-------TPSA        509
Thaps3_21900      YGLAIAPDNA-----------EGWKFPDVKTPLEGYYRCGDSTT--SGIG-------VPAT        574
Phatr2_45243      YGPRVEV--G-----------AGQTLPGHKTPLPGFYMVGDFTF--PGIG-------VPAT        565
Thaps3_11636      YGTQCVTEMF-------APINRNWTTTPFTEVPGLYLAGSDAFL-PSVTGAMYGGCLSAS        616
Phatr2_54800      YGTKCLPSMF-------AKSNRRWTTSPHTSIPGLYLAGSDAFL-PAVCGAMYGGCFGAI        877
Thaps3_5859       ------------------------------------------------------------        554
Thaps3_10254      YGLSHTPERF-------N---LQWL-KPKTPIQNFYLTGQDVCS-CGITGALVGGYLSAY        633

Thaps3_25361      AIMGYSFM---------DHMY--L-GKN-ITSDL-----------------QQFIEEPIL        955
Phatr2_42980      AVEQYGPL---------DHLF--L-QKN-ITTDI-----------------EQFLEEPGW        929
Thaps3_5221       ALNGASAANGF--VGIFDQWR-CM-DYLKAKGIIA*-------------------------        671
Phatr2_54826      AISGASAANAM--VSVFKQWE-CL-DELGKSQKL*--------------------------        635
Thaps3_21847      AASGTIAANTL--VDSSVQLD-LM-SELKDSGALQ*-------------------------        628
Phatr2_51868      AASGTIAANTM--TTIANQRN-LM-KELKAKGALQ*-------------------------        633
Thaps3_7094       AGSGMIAANSVSLDSIGAQLE-VL-SKIKQQ*-----------------------------        582
Phatr2_54842      AGSGLLAAHSVSWDSIGPQQD-LL-KTLQKRK*-----------------------------        595
Thaps3_10233      AASGAIAANTM--THVDNHLK-ML-SEASKLDPMYKFLDAGIMGQVYKPLVQGFTPSPEL        665
Phatr2_9210       AASGAIAANTM--NPVGKHLD-LL--------------------------------------        530
Thaps3_21900      ASSGAVCANAI--MSVWDQLS-LN-QKIKMP*-----------------------------        601
Phatr2_45243      AASGAIAANTL--VSVFDHLA-ML-DKVRLPEKEQKS*-----------------------        598
Thaps3_11636      AVLGTLGTMRLGH-AILTHLAMRLREENPKLSKIE----------AYMLAVKKFTE*---        661
Phatr2_54800      AVLGHLRALKLTL-AFIAHFAGCITDEDPKIGWIQ----------AYILAWKKFMND*--        923
Thaps3_5859       ------------------------------------------------------------        554
Thaps3_10254      AISPRCFLR--TA-SLLN*-------------------------------------------        648

Thaps3_25361      ATERNGVIVDDVAVPFKEVVVDMQKGITDADRSTAAESSKEE*      997
Phatr2_42980      VDEE------DVAIPYKSADAKKDKDV*---------------      950
Thaps3_5221       -----------------------------------------      671
Phatr2_54826      -----------------------------------------      635
Thaps3_21847      -----------------------------------------      628
Phatr2_51868      -----------------------------------------      633
Thaps3_7094       -----------------------------------------      582
Phatr2_54842      -----------------------------------------      595
Thaps3_10233      RTDQYVSGAGVAPVDYTATDPSVSERIDL*-------------      694
Phatr2_9210       -----------------------------------------      530
Thaps3_21900      -----------------------------------------      601
Phatr2_45243      -----------------------------------------      598
Thaps3_11636      -----------------------------------------      661
Phatr2_54800      -----------------------------------------      923
Thaps3_5859       -----------------------------------------      554
Thaps3_10254      -----------------------------------------      648
```

2) ZEP Percent Identity Matrix

1: Phatr2_47925	100.00	31.88	32.79	20.94	21.18	19.89	21.58	21.35	23.27	20.05	22.40	19.67
2: Thaps3_6395	31.88	100.00	46.63	22.93	18.67	18.94	21.71	26.03	24.36	20.46	21.14	19.61
3: Phatr2_45936	32.79	46.63	100.00	21.19	22.89	19.45	19.89	25.15	22.67	19.10	19.40	19.32
4: Thaps3_1961	20.94	22.93	21.19	100.00	22.54	22.92	24.21	25.13	24.02	22.81	26.11	22.77
5: Thaps3_270370_ZEP1_	21.18	18.67	22.89	22.54	100.00	65.42	33.85	33.48	33.67	20.00	20.93	21.04
6: Phatr2_45845_ZEP1_	19.89	18.94	19.45	22.92	65.42	100.00	33.41	35.44	35.52	19.56	22.22	22.65
7: Phatr2_56492_ZEP3_	21.58	21.71	19.89	24.21	33.85	33.41	100.00	44.88	42.80	20.00	21.03	22.19
8: Thaps3_261390_ZEP2_	21.35	26.03	25.15	25.13	33.48	35.44	44.88	100.00	78.51	21.62	23.16	25.08
9: Phatr2_56488_ZEP2_	23.27	24.36	22.67	24.02	33.67	35.52	42.80	78.51	100.00	21.76	23.33	23.84
10: Thaps3_22671	20.05	20.46	19.10	22.81	20.00	19.56	20.00	21.62	21.76	100.00	25.82	24.67
11: Thaps3_20663	22.40	21.14	19.40	26.11	20.93	22.22	21.03	23.16	23.33	25.82	100.00	44.80
12: Phatr2_43425	19.67	19.61	19.32	22.77	21.04	22.65	22.19	25.08	23.84	24.67	44.80	100.00

ZEP Alignment

```
Phatr2_47925            ------------------------------------------------------------ 0
Thaps3_6395             ------------------------------------------------------------ 0
Phatr2_45936            MADQSAARKTLPSLLRHFEGTELTVELKTGRLYRGTLSSADQAMNLTLEDASLLQRLIVN 60
Thaps3_1961             ------------------------------------------------------------ 0
Thaps3_270370(ZEP1)     ----------------------MTV---------RRIASLAIGISLSTLTCAFVTIS--- 26
Phatr2_45845(ZEP1)      ------------------------------------------MKFSTTVSSALFLIASV-- 17
Phatr2_56492(ZEP3)      ---------------------------------------------MK-RSCSIVTILY-- 12
Thaps3_261390(ZEP2)     ------------------------------------------------------------ 0
Phatr2_56488(ZEP2)      ------------------------------------------MGLSFL-SLCAVLTASS-- 16
Thaps3_22671            ------------------------------------------------------------ 0
Thaps3_20663            ------------------------------------MVSTILIFILVACLLLQST--- 19
Phatr2_43425            ------------------------------------------------------------ 0

Phatr2_47925            ------------------------------------------------------------ 0
Thaps3_6395             ------------------MSSSDRHSSQLNQPKRPRHEEPSHSI-----MAFDL--KN- 33
Phatr2_45936            QQHKGAFRRGSSSAAVPSTLSLVHIRGSTIRFIHFPDQLDLTLTIKQGIDRSWRMEHNSE 120
Thaps3_1961             ------------------------------------------------------------ 0
Thaps3_270370(ZEP1)     -----------SSRTTIKPLNVVGEQASSIGPATLLRNLKQNL-----PQIDWLAEGKGS 70
Phatr2_45845(ZEP1)      -----------STTTSFTPVQSFGVHR---------------------------------- 33
Phatr2_56492(ZEP3)      -----------VATT------------------------------------------- 16
Thaps3_261390(ZEP2)     ------------------------------------------------------------ 0
Phatr2_56488(ZEP2)      -----------AMAF------------------------------------------- 20
Thaps3_22671            ------------------------------------------------------------ 0
Thaps3_20663            -----------CDAFTF----------------------------------------- 25
Phatr2_43425            ------------------------------------------------------------ 0

Phatr2_47925            ------------------------------------------------------------ 0
Thaps3_6395             --RSPYEKKMERVITACPKCNGEGKVRAP---------LSKKARAQRKRMQQSQTGDTTN 82
Phatr2_45936            PGWDETTTIHEGPFTVCPKCHGDGHIVHQ---------ASKKQKLRHKRART--NGDYTD 169
Thaps3_1961             ----------------MPCSQLLSTAFNNDYNHLI--------------------TSQH 23
Thaps3_270370(ZEP1)     PSNK--IDIPDHVATVLAQPNAPKREAESEERTHKIRSRAKQASEDA--MALRGMLIG-D 125
Phatr2_45845(ZEP1)      --RT--LL----VTPRHATVEPPVREPETSDRVRQVRDRFRKASQDA--ANAKGCVAQDD 83
Phatr2_56492(ZEP3)      ------------VRAFAPAP--LVQ-----------SSCFFQRQPTT--TAR--FVSGTA 47
Thaps3_261390(ZEP2)     ---------------------------------------------MAD----DEAD 7
Phatr2_56488(ZEP2)      ------------VTTRSPACNDVTRSLH-RINTRHMTYPFYPASSLR--IST--RVASTA 63
Thaps3_22671            ---------------------------------------------MADSPTASSEE 11
Thaps3_20663            ------------------PSSGVLRDVRVSINGSVERRCSLPASEHVQHVAASSTSSSSS 67
Phatr2_43425            ------------------------------------------------------------ 0

Phatr2_47925            -------------------------MGKKRQSRPRLDPGAHIAIIGSGLAGLSTALSLE 34
Thaps3_6395             APNLAILKKPCKECDGSGLIAINPLDTTERKQTPPQIQPNFSVAIVGGGIGGIALAAALQ 142
Phatr2_45936            TPAP-QRLETCRECDSSGLVQSDT--------DPPVDTTLPEIAVVGGGLAGLALAAACR 220
Thaps3_1961             I----------------------NHI-MFIKQATLNHAKYASVAVVGGGIGGLTAANALL 60
```

```
Thaps3_270370(ZEP1)      DDAN-------------AWWR-EQRS-IPEGGRVVTTDDPLTVLVAGGGLAGLVVAAACH 170
Phatr2_45845(ZEP1)       GDES-------------SWWR-K--P-LPEDNDVISNQRPLRVVIAGGGVAGLVTAAACH 126
Phatr2_56492(ZEP3)       PPS--------------------SNVA-SEEKVDAISEAHRLKVLIAGAGVGGLSLAKVLT 87
Thaps3_261390(ZEP2)      ADF--------------------NSSD-YELLGRPARPGRPLKVAIAGGGVGGLTAALCML 47
Phatr2_56488(ZEP2)       VPP--------------------EDVA-FDKLSLPAREGRPLKIAIAGGGVGGLTTALCML 103
Thaps3_22671             VPS--------------------APPH-TTV--DDISNLEHHPLVIIGGGIGGLVLALCLD 49
Thaps3_20663             TRS--------------------RSST-TTLQAATAPHQPVQKVAIIGSGIAGLALAHAFL 107
Phatr2_43425             ------------------------------------------------------------ 0

Phatr2_47925             QG--------------------GFTNVHIYERDGSHDARKEGYGLTLTYNPTGVLHQLNV 74
Thaps3_6395              HR--------------------NIP-CIVYERDLSFEERKQGYGLTMQQGARALR-SLGF 180
Phatr2_45936             HR--------------------GMK-YTVYERDLDFHQRSQGLD---------------- 243
Thaps3_1961              NK----------------NPNLIER-LTVYEQAKEFTPT-AGAGFGFSPNGQICLSSIGI 102
Thaps3_270370(ZEP1)      S-----------------K---GMK-VALFEQASSYAPY-GGP-IQIQSNALRALQQINP 207
Phatr2_45845(ZEP1)       A-----------------K---GMQ-VAIFEQASQYAPY-GGP-IQIQSNALRALERINP 163
Phatr2_56492(ZEP3)       KM----------------P---TMD-VTVLEQTSEFKRF-GGP-IQLASNAMEILKHMDK 125
Thaps3_261390(ZEP2)      K-----------------K---GFD-VTVYEKTAAFARF-GGP-IQFASNALSVIKEIDE 84
Phatr2_56488(ZEP2)       K-----------------K---GFD-VTVYEKTAAFARF-GGP-IQFASNALSVLKEIDE 140
Thaps3_22671             QVYNHSITTDDANGEPITSSSVKFP-IHVYESTAEYSAN-AGGAIGLYPNGLRVLRNLSR 107
Thaps3_20663             S------------NNPSSSNNNKIQ-IDIFDSRTNLDEK-AGSGIQLT-GGLVALNEISN 152
Phatr2_43425             ------------------------------------------------------------ 0

Phatr2_47925             LE-----EIA------------------------QSDCPSRSHYM--FNA-NGEIQGYFG 102
Thaps3_6395              FSFSDDGEDD----NNNCSGKKA-VDENTSNTKQKFGIHSTRHVV--HKP-DGTVVGEWG 232
Phatr2_45936             ---------------------------------SDGIMSTKHVV--HEP-DGAIVGEWG 266
Thaps3_1961              YGYKKFILPFNSM--------KR-LNKEGNL----------------VNQSDV------- 130
Thaps3_270370(ZEP1)      EIFQELVTAGTCTADRVSG-LKIGYKKGNK--LA--------GL---YDAGDW------- 246
Phatr2_45845(ZEP1)       VICEEIRKAGTVTADRVSG-LKIGYKKGVFLGLG--------KQ---YEKGDW------- 204
Phatr2_56492(ZEP3)       PVFDKVMEKFTFTGDKENG-IKDGIRT------------------------EW------- 153
Thaps3_261390(ZEP2)      ELFERVMDKFTFTGTRACG-IKDGLRADGSFRMTNDSLDYLWNP---EAPADW------- 133
Phatr2_56488(ZEP2)       TLFERVMDKFTFTGTRTCG-IKDGLRADGSFRMTEDRLDYLWNP---DAPADW------- 189
Thaps3_22671             GSSPSYLDSEHVKFGANCPNLLQNVR--------TAGCDYIYRRWMRHDGLQVAVAREDE 159
Thaps3_20663             NLYNEVVESS-----------LPL------------ERLVSKCRPWFGGNKDDAGVEQGWQ 190
Phatr2_43425             ------MDAG-----------LLQ------------TGVRSRCKPWNPASPFDT------- 25
                                                                                 :

Phatr2_47925             NA-----------F-------------------------------------------A-- 106
Thaps3_6395              MK-----------V-----------------------------------WGGRFE-- 241
Phatr2_45936             LR-----------K-----------------------------------WGRSER-- 275
Thaps3_1961              LR-----------EL---------------------------------------------- 134
Thaps3_270370(ZEP1)      LV-----------RFDTIGP----------------------------------------- 255
Phatr2_45845(ZEP1)       LV-----------RFDTLQP----------------------------------------- 213
Phatr2_56492(ZEP3)       YA-----------KFDLKTP----------------------------------------- 162
Thaps3_261390(ZEP2)      FV-----------KFPLRQC----------------------------------------- 142
Phatr2_56488(ZEP2)       FV-----------KFPLKQC----------------------------------------- 198
Thaps3_22671             LLPDIKVDESEMAKLEVLDESEKGTGSRSSAVSRADSTKSQDVEGERANRRPHGGSFANA 219
Thaps3_20663             LL-----------ELDIQNA---------------------------------------- 199
Phatr2_43425             LL-----------DLDLLKT---------------------------------------- 34

Phatr2_47925             ---------------------------------------R-----NRGWGQRGNLRVPRQ 122
Thaps3_6395              ------------------------------------KN----GRKHAKRQNAHISRQ 258
Phatr2_45936             ---------------------------------------------AKKPKRQNIHIARQ 289
Thaps3_1961              ---------------------------------------------SNRHGFGIAGC-LRS 148
Thaps3_270370(ZEP1)      ---------------------------------------------ALEAGLPATVVVDRP 270
Phatr2_45845(ZEP1)       ---------------------------------------------ALDAGLYPTVVVDRP 228
Phatr2_56492(ZEP3)       ---------------------------------------------AENRNMPYTGVIERP 177
Thaps3_261390(ZEP2)      ---------------------------------------------ADLFGLPYTGVIDRP 157
Phatr2_56488(ZEP2)       ---------------------------------------------ADLFGLPYTGVIDRP 213
Thaps3_22671             MGALEAMKDMSQNLSQRLSRISFTGSDATTTSAAGGSDKSTPRASRVVDTELLSLGIRRW 279
Thaps3_20663             ----------IRENA---------AADASKQHGAEEGDSNKQY--SLVREDGEVVAYTILRG 240
Phatr2_43425             ----------VQNA----------------------GSDV-S--NALIREGKLVWTSIMRG 60
                                                                              *

Phatr2_47925             RVRQILASRL---KITETHWDHKLVGVSGCENGE-------NICLAFQLEGA-AEEKLLV 171
Thaps3_6395              NLRQLLMEML---HPGTIQWGQKFVGYSGQSSDDDSSQDQPSLQVRFRRRSNDCDEEVAT 315
Phatr2_45936             SLRWQLYKAA-GGRTANIAWNHRLLQYKQRV-------DAPGWELKFQV----DDQIIAH 337
Thaps3_1961              DLVNLLVEQL-----DTQHGGKGALKYSEKLVGINPIH--DKVELEF------ESGRQD- 194
Thaps3_270370(ZEP1)      VIQQILVKYG--FPEGTVRIKSRIQSYEDL-----GKG--RGVSVTL------EDGTKA- 314
```

```
Phatr2_45845(ZEP1)       VIQQILLEHG--IPEKTVRIKSRIANYEEL-----GPG--KGVRILL------EDGTVA- 272
Phatr2_56492(ZEP3)       DLQQIFLDSLPK---GTVKNGDGVARYEKL-----PDG---GVKAVL------KSGKEV- 219
Thaps3_261390(ZEP2)      DLQEILLDECRKIKPDFIQNGNPVNGYVSK-----GKG--NGVTVNL------ADGTTA- 203
Phatr2_56488(ZEP2)       DLQEILIDECRKLKPDFLINGNPVVGYEDL-----GKG--QGVTINL------NDQTTA- 259
Thaps3_22671             KYQQVLYDQC-KEVGIQFHMGKRLQSVTSIPASGEDGD--AKSLLLF------KDGSRI- 329
Thaps3_20663             TLQRILREQLAQEHGVDVQFDKRLCGMAY---SNEENG----VKCQF------NDGTTTG 287
Phatr2_43425             ALQEALYGALPSNVRQNVQFGKVLVDLR----SVREGG----IECLF------SDGSVAG 106
                                :                                        :       .

Phatr2_47925             GADLVVAADGIRSAVLQHAYP-QAPPI---------------------------------- 197
Thaps3_6395              TASVLVGCDGIRSSVRSAKLGEDGTPL---------------------------------- 342
Phatr2_45936             KADLIVGADGLRSQVRRSLIGEDRTPL---------------------------------- 364
Thaps3_1961              LVDLVIGADGINSSVSKLLNIDDEIA----------------------------------P 221
Thaps3_270370(ZEP1)      YADVLVGADGIWSQVRKNLHGLDDGAGGFAASGAAGGALDDAEARKLARDTVAIAAKADR 374
Phatr2_45845(ZEP1)       YADVLIGSDGIWSSVRRIMHGLDQGADGFAASGAAGGALNEAEARRMAKDSVLMANNANR 332
Phatr2_56492(ZEP3)       YGDVLIGADGIWSAVRATMRDS---------------------------PAKGDGSGA 250
Thaps3_261390(ZEP2)      EADVLVGSDGIWSAIRAQMYGEEI-KKS----------------------SNNALKRQGC 240
Phatr2_56488(ZEP2)       SADVLVGSDGIWSAVRDQMYKEGGVKST----------------------SANKKKRQGC 297
Thaps3_22671             TASLVIGADGINSKVRNYVTNPKPTAAT-----------------------TKQQEEYVP 366
Thaps3_20663             PYDLVVGCDGIQSKVKQYVNTGSLQPNA---------------------------DSSSA 320
Phatr2_43425             PFDVVVGCDGIKSACKEYVENGRILPKD----------------------AKREGDSVA 143
                           .::::.**: *

Phatr2_47925             QSLGIRLILGISSSF----------------THVHLKER--GFYTLDSGKRLFVMPFART 239
Thaps3_6395              RYLDCIVILGIAPSP---------------TSALTDGET--VFQTADGITRLYVMPFAEA 385
Phatr2_45936             RFLDCIVILGICPIAGISL--------EGQQSDLLDGET--VFQTADGVTRIYIMPFTTT 414
Thaps3_1961              IYSGANIFYGKIPNPDGHE--------YLRGHPIFTEG------SVTNGPGTGEFI---- 263
Thaps3_270370(ZEP1)      RFSGFTCYAALAPHRASNI--------ENVSYQILLGEK-KYFVSTDGGGDRQQWF---- 421
Phatr2_45845(ZEP1)       RYSKFTCYAALTEHRASNI--------EEVSYQILLGKD-KYFVSTDGGGERQQWF---- 379
Phatr2_56492(ZEP3)       TYSGYTVFAGELAYDSFDN--------GQVGYKVYIGPG-QYFVITDIGNGNYQWY---- 297
Thaps3_261390(ZEP2)      TYSGYTVFAGETVLKTEDY--------YETGYKVYIGPQ-RYFVTSDVGDGRVQWY---- 287
Phatr2_56488(ZEP2)       DYSGYTVFAGETILKTPDY--------YATGYKVYIGPK-RYFVTSDVGDGRIQWY---- 344
Thaps3_22671             AYTGVTCLMGCASVPRIRGICFPSSAT-TKCHACYYPTRAPKEVDDEGNADDTVRPVSGD 425
Thaps3_20663             IYSGIRITFAIQEGDADDN---PVQAKKGAQFTQFFGNG-AYALTSSYGAGKGVPPAKG- 375
Phatr2_43425             VYSGLRIRYAVKDGNSEEK---QA---ETATLSQYFGEG-AYGLDGIYGAGPGQPHTKC- 195
                                  .

Phatr2_47925             AEWAVIDDNATSCSQPTGEQYMWQLSFASSDDKTY-----SSTEL----LQQALFHCRDW 290
Thaps3_6395              GD-----DSSGLSTDNTKGLSMWQLSFPMDETDATRLSQLGSSAL----KEEALKRCGAW 436
Phatr2_45936             -------------------AYMWQLSFPMAEEQALSLSSKGPSAL----KKEAIRRCQSW 451
Thaps3_1961              ---AF----HTGAEDN--KTFIWANTY--ASNSPPPK-------------------REDW 293
Thaps3_270370(ZEP1)      ---AL----IREPAGGVDPEPT--------PEDPHPKLTRLRKEFACNGS--GDADGNVW 464
Phatr2_45845(ZEP1)       ---AL----IREPAGGVDPEPT--------PENPTPKLTRLLQEFNHEEP--GDQNGDVW 422
Phatr2_56492(ZEP3)       ---AF----LARPADSASST------------DMPDGQSKYLQEI-----------FAGW 327
Thaps3_261390(ZEP2)      ---AF----FALPPGTKKAPSGWG---------------DYIKSL-----------HQGW 314
Phatr2_56488(ZEP2)       ---AF----FALPPGTKKAPSGWGGSTRDGQTDPEENLVDYVKGL-----------HEGW 386
Thaps3_22671             YEQVF----QIYFPSPIERPDTWRTLTP-------TEAKEECRELA------KKLREDGW 468
Thaps3_20663             ---AF----LIYTDDDYYGP-----FPK-------SVFKRDESEVAVKPSVEAAAENADW 416
Phatr2_43425             ---AF----LVYLDPDYVGP-----FKK-------KRAR-----LESKPS---VDENADW 228
                                                                                   *

Phatr2_47925             HEPVQE----------------------LMLS-TSQESIWGTLLYDRNP---EILHKH-- 322
Thaps3_6395              HDPILK----------------------LLRS-TPEDFITGYPCYDRALVERKELRDG-- 471
Phatr2_45936             HTPVPD----------------------ILYS-TPIELVSGYPVYDRALLTPELLQE--- 485
Thaps3_1961              SEGNFHELKDILLKYPTS-----HPIHKFAEL-TGESDLLHFGLYYR----------HHK 337
Thaps3_270370(ZEP1)      DPFALE----------------------LINA-ASEEDIKRRDLYDGAPLLTTLDPQRLL 501
Phatr2_45845(ZEP1)       DDFAYE----------------------LFKA-TPEEDIKRRDLYDGSPLL--------M 451
Phatr2_56492(ZEP3)       SEEVHH----------------------ILRA-TQEHEIEQRDLYDRPPSA--------M 356
Thaps3_261390(ZEP2)      SDEVMT----------------------VLDS-TPPDSVEQRDLYDRPPEL--------L 343
Phatr2_56488(ZEP2)       SDEVMM----------------------VLDS-TSPDSVEQRDLYDRAPEL--------F 415
Thaps3_22671             DEQFLAPLESETL--TGVL-RVGLRSRE------------ALDV-WHVGG---------S 503
Thaps3_20663             TQDNRVPREHVAE-CIKVLKTAAIPGNDVADIVSNSNRFFDLGVYFHNPFS--------W 467
Phatr2_43425             TQDVRKSIEVARETMLDQTKSLGVPDTDLSPTISAADRFFELGVYFHNPFS--------T 280


Phatr2_47925             ----LQINDTLPRRILIVGDACHAMSPFKGQGANQALQDGRVLVKHLTSARVE------- 371
Thaps3_6395              ----CDKSQSANAFVTLLGDACHPMSPFKGQGANQALLDAVLLSQKLFDISRIHNGKTNV 527
Phatr2_45936             -----------TASVTLVGDACHPMSPFKGQGANQALLDALALVRSIYKHCKTAGSY--- 531
Thaps3_1961              NTWS-------KDRVVLLGDACHATLPYVGQGANQAIEDAIYLAVCLNR----------- 379
Thaps3_270370(ZEP1)      SPWA-------KGPVALCGDAAHPMMPNLGQGGCQATEDGYRLVEELAKVQH-------- 546
Phatr2_45845(ZEP1)       QGWS-------KGQVAICGDAAHPMMPNLGQGGCQATEDGYRLAEELATVRT-------- 496
```

```
Phatr2_56492(ZEP3)       KPWT-------DGPVALLGDGVHAMMPNLGQGGCQAIEDAFVIGQELGSATK-------- 401
Thaps3_261390(ZEP2)      RSWA-------DGNVVLIGDAVHPMMPNLGQGGCQAIEDAFVLSETLEACES-------- 388
Phatr2_56488(ZEP2)       RSWA-------NGNVVLIGDAVHAMMPNLGQGGCQAIEDAYVLTETLANTRT-------- 460
Thaps3_22671             IGVASGEDNDDVGRAVLLGDAAHPPVPYIGQGAMMAMEDAGTLALLLARYCPLDTTNSPT 563
Thaps3_20663             NGWVREFDKSA-KYAVLAGDAAHAMPPFLGQGANQALQDAYLLAEKVFEYNDQVEQYSPV 526
Phatr2_43425             QGWSREMTDSKGGSVVLCGDAAHALPPFLGQGSNQAIQDAYCLAKQLYAYNAEIEQG--- 337
                                       : **. *   *  ***.  *  *.  :   :

Phatr2_47925             --------------IAVSNTQREIVQRTASVVAASRQASVYWHEPQLVMPKDGQDTQKFA 417
Thaps3_6395              NEQQPTISLNESTPQALAEFENDMLQRCEVKVKKSADAAKFLHSDVA------------- 574
Phatr2_45936             --------DSNSLERAVKEFEVEMLVRSAVKVEASAEAARFLHTEIA------------- 570
Thaps3_1961              ---------HDNYSDAFADYYDKRFPRTKRIVQFAGIMHKLYHT---------------- 414
Thaps3_270370(ZEP1)      ---------SRDVPGALGRYSRVRVIRTAIIQGFAQLGSDLLVD--------------FD 583
Phatr2_45845(ZEP1)       ---------TKDIEGALQEYYRKRIPRTTIIQALAQLGSDLLVD--------------FD 533
Phatr2_56492(ZEP3)       ---------RSQIVDKLREYQQRRLIRSAAVQGLSRFASDIIIR-------------GFD 439
Thaps3_261390(ZEP2)      ---------TQKLEDALQDFYKKRIVRVSIVQFLSRLASDLIIN-------------AFD 426
Phatr2_56488(ZEP2)       ---------TEKLQDALQEYYRKRIVRVSIVQFLSKLASDLIIN-------------AFD 498
Thaps3_22671             VD-------FSLFKKAMHAYESLRVSRTKTILGSSVELGKTQQKRAE----S-----KLY 607
Thaps3_20663             VRGGE-STAEPNLKALLNEYEKRRWLPTTSITAKAAGLGYLETG-S-----------GFF 573
Phatr2_43425             --------RDANLNAMLKDYENTRWPSTFGIFWKSTFLGYLETG-GE----D-----GLY 379
                                        .               :

Phatr2_47925             GVCSQDIPALLHALQKKNIKANSANDLDKSVQCTI---DELQLIGPPEVRRKAESELKEA 474
Thaps3_6395              -------------IQEGNITRGAAA-LD--------------AKG*-------------- 591
Phatr2_45936             -------------IQKGNVTRGAAS-RS--------------ILIKSNEEDTATTE*--- 598
Thaps3_1961              DS--------------WLVHKA----LDVLIGSIINGGAALKQLEREIINECPVKDYQQY 456
Thaps3_270370(ZEP1)      LM--------------MTI------------------------------------------ 588
Phatr2_45845(ZEP1)       KM--------------MTI------------------------------------------ 538
Phatr2_56492(ZEP3)       TP--------------AKIYRD----ENGK----------FQ------FENCNYAGIVTK 465
Thaps3_261390(ZEP2)      TP--------------WSPHDD----LGKS----------WK------S-------YLTF 445
Phatr2_56488(ZEP2)       TP--------------WSPHDN----LGKS----------WK------S-------YLTF 517
Thaps3_22671             NA-----------WREWSIKA-----------------QVWAYGTLPVMRPGAASDYMTK 639
Thaps3_20663             GN-----------FR*------------------------------------------- 577
Phatr2_43425             AR-----------FRDVFFKT-----MGAV------GIAEWVLI--SAAKPKI*------ 408

Phatr2_47925             HFVDLARQAILSSEMNNLAFLRKLSWEYPNLIR-----------NV-DVDDMTCLQKAAK 522
Thaps3_6395              ------------------------------------------------------------ 591
Phatr2_45936             ------------------------------------------------------------ 598
Thaps3_1961              AFNRQ*--------------------------------------------------- 461
Thaps3_270370(ZEP1)      ---PLLGPFFLTMTQLSMPFILRYLYTPSF*------------------------------ 615
Phatr2_45845(ZEP1)       ---PLVGPFFLFMTQVSMPFVLRFLYTPEF*------------------------------ 565
Phatr2_56492(ZEP3)       ILQP-----ILPI-FFSVQFAFLYDG----------------WKNDKQIDFKAFLGFSVL 503
Thaps3_261390(ZEP2)      FWKVRGVAPFYGNECQHISYSLLFSTQQPS------------------------------ 475
Phatr2_56488(ZEP2)       FWKPILQFAIFPM-----QFAYLYSYYPTGNMGDLPAKLEAIWKEKHKTDAEAVFEQASK 572
Thaps3_22671             VEEVVLE*------------------------------------------------- 646
Thaps3_20663             ------------------------------------------------------------ 577
Phatr2_43425             ------------------------------------------------------------ 408

Phatr2_47925             SGHIRIA---HWLITEAGCL--VDA-RLL---NDLSIKSYMKALLKMYM*-------    562
Thaps3_6395              ------------------------------------------------------------    591
Phatr2_45936             ------------------------------------------------------------    598
Thaps3_1961              ------------------------------------------------------------    461
Thaps3_270370(ZEP1)      ------------------------------------------------------------    615
Phatr2_45845(ZEP1)       ------------------------------------------------------------    565
Phatr2_56492(ZEP3)       GGLIVSL--VLFELAEAGLGIGLGAEALLGAEGLLDFGGISAAIQDFFLGGGAAGL*    557
Thaps3_261390(ZEP2)      ------------------------------------------------------------    475
Phatr2_56488(ZEP2)       EGFVMEHEASFFKKAEVE----LSPTALAATKEELS*--------------------    604
Thaps3_22671             ------------------------------------------------------------    646
Thaps3_20663             ------------------------------------------------------------    577
Phatr2_43425             ------------------------------------------------------------    408
```

3) VDE/VDL/VDR Percent Identity Matrix

1: Phatr2_56450_VDR_	100.00	52.25	17.27	17.45	14.40	17.05	16.82	16.67	13.54
2: Thaps3_270211_VDR_	52.25	100.00	16.24	17.33	16.25	16.05	18.10	16.75	11.46
3: Thaps3_7677_VDE_	17.27	16.24	100.00	56.04	26.69	23.60	21.62	22.31	18.89
4: Phatr2_44635_VDE_	17.45	17.33	56.04	100.00	26.49	24.01	21.93	21.65	15.56
5: Thaps3_22076_VDL1_	14.40	16.25	26.69	26.49	100.00	65.61	30.81	30.30	30.21
6: Phatr2_46155_VDL1_	17.05	16.05	23.60	24.01	65.61	100.00	28.57	28.30	28.42
7: Thaps3_11707	16.82	18.10	21.62	21.93	30.81	28.57	100.00	53.25	39.29
8: Phatr2_45846_VDL2_	16.67	16.75	22.31	21.65	30.30	28.30	53.25	100.00	54.46
9: Phatr2_bd_1281	13.54	11.46	18.89	15.56	30.21	28.42	39.29	54.46	100.00

VDE/VDL/VDR Alignment

```
Phatr2_56450(VDR)       MKLHRKGRYRLLVTAVLLGTVCSFVPENLRSGSVRIPRKNANAGSVPGTHTVSKQPSASA  60
Thaps3_270211(VDR)      ------------MAMVLLIRTAV-----IASYSLTL----------TSAFSSSIRPTCRT  33
Thaps3_7677(VDE)        ------------------------------------------------------------  0
Phatr2_44635(VDE)       ------------------------------------------------------------  0
Thaps3_22076(VDL1)      ------------------------------------------------------------  0
Phatr2_46155(VDL1)      ------------------------------------------------------------  0
Thaps3_11707            ------------------------------------------------MSASSSS-------  7
Phatr2_45846(VDL2)      ------------------------------------------------MK------------  2
Phatr2_bd_1281          ------------------------------------------------------------  0

Phatr2_56450(VDR)       TRHKVSQTSIDANALPIKNDLIQGPVSANKSIGSITFLLPSSGADEIKTNFGSSSPVGNP  120
Thaps3_270211(VDR)      FRQS--------------TPHHATASNVDIIGTVALLVPSSSTE--LSKYGSKSPAPRP  76
Thaps3_7677(VDE)        ----------------------------------MK--------LFL-----SLVLAAAP  13
Phatr2_44635(VDE)       ----------------------------------MKFLGVTSLLCLW-----SVVNRENV  21
Thaps3_22076(VDL1)      ----------------------------------MRP-----STSAL-----TVVLGT--  14
Phatr2_46155(VDL1)      ----------------------------------MRFAWVVAAGVVL-----TTTTQA--  19
Thaps3_11707            -------------------------TTTTTNAGKRARSWPSSSASTS-----SMPTRS--  35
Phatr2_45846(VDL2)      -------------------------RATRKRTLAATLWIAMSSVTG-----SGPGRT--  29
Phatr2_bd_1281          ------------------------------------------------------------  0

Phatr2_56450(VDR)       SLDEAVRHLANKSQYFSDGRVETRIVYVPIEEQDESVWKDLLETDVLLAMGLQYEADLAF  180
Thaps3_270211(VDR)      SYQEAAEHLARKISHFSDGRIEATVVTPSTNQDDT--DDVCLTSNALIALGITDPAEVQY  134
Thaps3_7677(VDE)        VSSFAPSNPVVSR---THSSVHSQQHN-H--------------------VLEAHND-----  45
Phatr2_44635(VDE)       SEAFAPRHQSLSR---PSSRTTSAFSRAP--------------------ILSLRK------  53
Thaps3_22076(VDL1)      -IALV------SCSQ-LN-----------------------------------NNVS-AF  30
Phatr2_46155(VDL1)      -LVPL------DCTG-MGETRTSGI--RP--------------------IRGLESNMA-RY  49
Thaps3_11707            -ILILATFLSLTSSS-SSTSVEAAFVGSP--------------------AVGLRSHTAAST  74
Phatr2_45846(VDL2)      -AAFAP----------SGNNNNNGCHGLS--------------------RVALHTTELQAH  59
Phatr2_bd_1281          ------------------------------------------------------------  0

Phatr2_56450(VDR)       A--RKLFQQRH-DRDAEHRFRQCHFAIDCAQ-SFPTMVGPYDSENPSFRAKLLPWTKHAS  236
Thaps3_270211(VDR)      L--STTFRKRRTSHQETSSYNTCQFALDCGSNNYAPLVGPWDEANPSILAEIAPWTGVAS  192
Thaps3_7677(VDE)        -----------NMDDITFSLSARNINNE-------IVE-R--------IGKV--------  70
Phatr2_44635(VDE)       ------------------YDSDSEVEND-------LLS-K--------LNPFQNW--QTA  77
Thaps3_22076(VDL1)      STRSSSLTQRHKTCTITTS-SSSLYVNP-------N-NDD--------DNSNRSQHKPNP  73
Phatr2_46155(VDL1)      ATVRHGTDQ--TNHGITSS-SERQWPFP-------R-GG-----------------SSPRA  82
Thaps3_11707            SKQQSSLYAQ-KKNNIDSSDNPLSYLFD-------LSSDP--------ETKRRLQKQTAT  118
Phatr2_45846(VDL2)      SKPPHQQSPQ-T-----RYTPPSAMNIQ-------F-PEP--------DESLHVWDHVRN  97
Phatr2_bd_1281          ------------------------------------------------------------  0

Phatr2_56450(VDR)       GKRLSQQMVALLQRGNSDDF----VFAIMLF---LNQFSGSSVDWVKHSIDATWEKGPLR  289
Thaps3_270211(VDR)      GKRLTEQMNGLFEKQTSDEF----ALAVMLF---FNRFSGAAIPWVQHSIDVTWEKGLVQ  245
Thaps3_7677(VDE)        T----TSALLALTLSFSAI----------------TSPISGPNGDVLSSIPSAN----AAD  107
Phatr2_44635(VDE)       L----QSTALALTIGVASW---------------TSL------PTIVPPAFAA----TTD  108
Thaps3_22076(VDL1)      F----LSAALTAAVTTSLFLSSLPS--A-------TFASTPASTTQKYDGFAEYAKENKM  120
Phatr2_46155(VDL1)      V----ARSVATFGLGFSIALASVFGVAA-------APVGADTTPAVKYDGFAEYAQDNQM  131
Thaps3_11707            ---------LFSTLGFSALFLTNPLIPHLPFSPSLSSANAEDELYAKYGGK-----GLDT  164
Phatr2_45846(VDL2)      --------IGKVCTGFAL---AGLLSALVSF---TSPVWAENELSAKYGG------GLDT  137
Phatr2_bd_1281          ------------------------------------------------------------  0
```

```
Phatr2_56450(VDR)       NAQEVVSMVSKCGDCVVKCV-QDDNCRECLEVLTALDTR--DQVASYRTIVSYESDLLKD  346
Thaps3_270211(VDR)      NAKEIFSMITKCGPCITKCL-NDENCSQCINALDKIDTR--DQVTSYRTVVSFESELLRD  302
Thaps3_7677(VDE)        GAKIGLCLVKKCRVPLAKCI-TNPNCLANVICINSCNGKEDETGCQINCGNVFENDVVGE  166
Phatr2_44635(VDE)       SKSIVSCLFQKCPLPLAKCI-ANPKCLANVVCINTCTGRPDEIECQIECGNLFENEVVGE  167
Thaps3_22076(VDL1)      EQSDVGCFINKCGDQTKQLF-SNPRGIKGVSCLGRCKGE---QSCATRCFAEFGSEDLDN  176
Phatr2_46155(VDL1)      EQSDVGCFINKCGDQTKALF-SNPRGIKGVSCLGRCKGE---QSCATRCFAEFGSESLNA  187
Thaps3_11707            SLVDKDCLVNQCQVQAKACLQDDPDCRKGLTCTAKCLGD---NACITGCFARYGNENLDE  221
Phatr2_45846(VDL2)      SLVDQNCLVSACSLQTKACLQDDPSCRKGLTCTAKCLGD---NACITGCMARYGNANLDN  194
Phatr2_bd_1281          ----------------------------------------------MERYSNDKLNN  11
                                                                       : .  :

Phatr2_56450(VDR)       FSFCILQKNNIFNCDASLPTLPNVQPVATWREQPLTEDIARSLLVGHLNDEA-APESSLR  405
Thaps3_270211(VDR)      FSLCILQKNNIFECSAEIPELPVVKPMSTWRGKDVTTDVARGIMIGHLEGAGGSLEGNLQ  362
Thaps3_7677(VDE)        FNKCAVTDMTCVPQKKDDGSYP---------------VPSKDVLVQSF--------DTKL  203
Phatr2_44635(VDE)       FNKCVLTDMKCVPKQGDDGSYP---------------VPAPEIVVPKF--------DTKF  204
Thaps3_22076(VDL1)      WLSCTIEDYECVKVPKNIDN----S----------AENVGYDTTVKKF--------DPST  214
Phatr2_46155(VDL1)      WLSCTIEENECVKVPKNVDN----S----------AEDIGYSTTLRSF--------DPQS  225
Thaps3_11707            LLKCTIEDHECIKVAILEGGGDVLG----------REPKSPAPTVQGF--------DLAS  263
Phatr2_45846(VDL2)      LLKCTIEDHECIKVAILEGGADVFG----------QEPRAPAPTVTAF--------DPKS  236
Phatr2_bd_1281          LLKGTIEDHECTKVAILEGGIEVFG----------QELRASDSTVTAF--------DPKS  53
                             : .                                .    :  :        .

Phatr2_56450(VDR)       TDISWKVACGANEAYDKFPSQNQLFYPAARGRD---------LW----------------  440
Thaps3_270211(VDR)      LGVSWKVACGANVAYDQFPSQNQLFYPSAKGKD---------LW----------------  397
Thaps3_7677(VDE)        WNGRWFITAGQNKLFDTFPCQVHFFTETAPGKF----------VGK--------------  239
Phatr2_44635(VDE)       FDGRLYISAGQNKLFDVFPCQVHFFTETEKGKF----------FGK--------------  240
Thaps3_22076(VDL1)      LVGKWYKTDGLNPNYDLFDCQSNTFDFSDDTK-------------------------KELD  250
Phatr2_46155(VDL1)      LVGTWYKTDGLNPNYDLFDCQKNTFTP--TSD--------------------------KELD  259
Thaps3_11707            MEGTWYKVAGYNPNYDCYACQRNTFSSPEGGLSDSLQLPTGGILGSLSNAVGSIGADRLQ  323
Phatr2_45846(VDL2)      LQGSWFKVVGYNPNYDCYACQRNTFSAPDSSANGNRNN---LLWSVASGNTNPAVTNQLR  293
Phatr2_bd_1281          LQSSWFGKVAHNPNYDCYACRGIQSMDV-------------LIVFHCGETK----DRDR  95
                                . *  :* : .:

Phatr2_56450(VDR)       YDPVFRVETL--DGR--------------------------NVWCKRH------YKVRPA  466
Thaps3_270211(VDR)      YDPVFRVETI--DGR--------------------------NVWCKRH------YKVRNG  423
Thaps3_7677(VDE)        --LNWRIE--EPDGE--------------------------FFTRD----AVQEFVQDP  264
Phatr2_44635(VDE)       --LNWRVE--QPDGN--------------------------FFTRD----ALQEFVQDP  265
Thaps3_22076(VDL1)      MGIFFRVPRPEEYGG--------------------------GFWENSLTEHMIVDAVSP-  283
Phatr2_46155(VDL1)      MGIFFRVQRPPESGG--------------------------GYWENALTEHMIVDVPVQP  293
Thaps3_11707            VDVEFSMPRYLPDGSPQPPSGVRESFISSADSMEGSGLQSVGYNQYSTHETMVFDTVKSN  383
Phatr2_45846(VDL2)      MDVEFSMPHLLPDGSPPPPSNVRESILVSGEDGSVFGSKSIALNDYRTRETMVFDQVSTG  353
Phatr2_bd_1281          KGINFFLTHFFDADTVQ*-------------------------------------------  112
                            : :      .

Phatr2_56450(VDR)       DIPG-TFRF----------------------SVLDNGITSNEFWTIVGVADD-----LSW  498
Thaps3_270211(VDR)      ETPG-TFKF----------------------SVLDNGVTSNEFWTIVGAADD-----LSW  455
Thaps3_7677(VDE)        NNPA----------HL--INH------------DNEYLHYQDDWYIVDYAADDNKEGVPP  300
Phatr2_44635(VDE)       NQPG----------HL--INH------------DNEYLHYEDDWWVIDYEYDGNKDGVPP  301
Thaps3_22076(VDL1)      -------------------ELDNPTGRTMHTAGKMYGLKFTENWYILGE--SNGDNDIPP  322
Phatr2_46155(VDL1)      PTAGTQLVASANAATGDLNDELNPTGRTMHTAGKMYGLEFTENWYILGE--SDGKGSVPP  351
Thaps3_11707            GVGE-AVKL-----ALGKRGEEKLYSRTAHSEGEMFGLKFWENWYIIGQ--NNP--GQDE  433
Phatr2_45846(VDL2)      N----NMVF-----H-KGTTQEVSYSRTAHSEGEMFGLKFWENWYIIGE--NDP--GQPE  399
Phatr2_bd_1281          ------------------------------------------------------------  112


Phatr2_56450(VDR)       IVFHYAGAASAVGQRYLGGLLCTADGSLPDESQRPEIWRVLRSAGIQPWDLYT-VNNDLT  557
Thaps3_270211(VDR)      VVFHYAGAAGAVGQRYLGGLLCTPTGELPPEEDLGHIYNFLRSAEIEPWELFV-VDNDDQ  514
Thaps3_7677(VDE)        FAFVYYRGENDAWIGYGGAVVYTRDSKLPE-SLLPRLREAAKKVNFDFDKDFDLTDNSCK  359
Phatr2_44635(VDE)       FAFVYYRGNNDAWEGYGGVVVYTRAAQLPE-SLLPRLRVAAEKIGFDFDKDFVITDNTCP  360
Thaps3_22076(VDL1)      FKLVAYKGHTL-QGNYEEAFVYAKESVLPK-EAVGAVREAAAKAGLDFDK-FTRIDNTCP  379
Phatr2_46155(VDL1)      FKLVAYKGHTL-QGNYEGAFVYAKEATVPE-AAKPAIREAATKAGLDFDA-FTRIDNTCS  408
Thaps3_11707            FKFVYYNGKTR-QNTYDGAFIYSRSRTLSP-ASMEKVYKIAKDAGMNPDQ-FCKIQNSCF  490
Phatr2_45846(VDL2)      FKFVYYNGKTR-QNTYEGAFVYSRSKELAP-ESMAKVYSIAKEAGMKVDQ-FCRIRNGCF  456
Phatr2_bd_1281          ------------------------------------------------------------  112


Phatr2_56450(VDR)       SPG------------AQEAGPPPLDYFRREVLAKR---------------AASSPHP*--  587
Thaps3_270211(VDR)      SPG------------ALAAGAPPLDYFRKTASVIG---------------*---------  537
Thaps3_7677(VDE)        ALEKG-EEVVLR-EKFAGKMAIQTEK-------------QLQQQAVLARTAASNTVKGEV  404
```

```
Phatr2_44635(VDE)       TDLSGKEKQILR-EKFAGKVALQTEQ-------------QLQAQVTRLRGNAVNSIKAQK  406
Thaps3_22076(VDL1)      TTTKS------LNDASAGTG-TSTTDWVDLVVGEGGVIDWV---VPGWRGEYKN*-----  423
Phatr2_46155(VDL1)      V-GDS------LNDAQAGTG-TSTTDWINLVVGEGGVIDWI---SPGWRGEYKAKR*---  453
Thaps3_11707            DGEDDKQEMMMMNPQREGLG-SPSNPFRGILASTK-VSQFLGVESVAAETTYNEPKSTIS  548
Phatr2_45846(VDL2)      SDET------VVKAPSSGLG-SQSNPFRGILASTR-ISQLLGVEPVAARDTVRRNAPT--  506
Phatr2_bd_1281          ------------------------------------------------------------  112

Phatr2_56450(VDR)       ------------------------------------------------------------  587
Thaps3_270211(VDR)      ------------------------------------------------------------  537
Thaps3_7677(VDE)        TAVEKSLQK-------IEEKALAFEKELMKDVVSVEKEIVKEVEEVEKEIVQEEQKIFGG  457
Phatr2_44635(VDE)       LFFEQGLEGAQKAYDALEETEKQFERETSQQ*----------------------------  437
Thaps3_22076(VDL1)      ------------------------------------------------------------  423
Phatr2_46155(VDL1)      ------------------------------------------------------------  453
Thaps3_11707            SNFLQGSQATNRKADAVQERPWWKEMGDY-------------LEDP-----RRHFRLMDS  590
Phatr2_45846(VDL2)      ------SPTLQPAVGTIASRPWWYEIGDY-------------LENP-----HRHFQVMDS  542
Phatr2_bd_1281          ------------------------------------------------------------  112

Phatr2_56450(VDR)       -----------------       587
Thaps3_270211(VDR)      -----------------       537
Thaps3_7677(VDE)        IR*--------------       459
Phatr2_44635(VDE)       -----------------       437
Thaps3_22076(VDL1)      -----------------       423
Phatr2_46155(VDL1)      -----------------       453
Thaps3_11707            LRTDMDWPDYIKEKNW*       606
Phatr2_45846(VDL2)      LRLPMTWAEDVKN*---       555
Phatr2_bd_1281          -----------------       112
```

APPENDIX 2.C FULL-LENGTH GENE MODELS, PROTEIN TARGETING PREDICTIONS

1)Thaps3_268908 (PSY1)

Exon 1:
>Thaps3 chr_5:1332233-1330996
TACTTTTGAAGCCTTCTCCCTACTACTTCTCAATACGTTTTGCAATCTTTGCTGCATAAA
CATTAAAGATTGATTTTCGCCGGTTGATACACAACGATACGGAAGAAGTGAAAATCAATT
GCAAAACGGAACAGTACCTCTGCCGTAGAGGAACACAACATGAGGATATCATCCATTGTA
GCAGCCACCCTGATCACGTGCGATGTTGCTTCAGCATGGTCATCGTCTGCATTCACATCA
CCGTTATCTATCAAGACACAGCATCGGTCTCAGGCATCAACCAAAAGAAGAGACGGATCA
ATCTTGATGTCTGCTGTGTCACAAAAAGAAAGTAGTAGTAGTAGCAGCAGCGGCCAAGGA
CGCACCACAAAAGAGATTTCAAACGAATTAGCATTAGGTGTTACTCTCGATGGAGGTCCG
GTGATTGATTTTGCCTCCGTCAAAGATACCACATCTCGTGCAGAAATAGCACTTGCCGAC
TCACGAAAGAAGTACGAAGCCAGTGGTGCTACAATTAGCCCTAATCCGGGAGGAAGATTA
ATGGGTATCAACGATGAGGTTGTGGCAGAGGTTGGCTATGAGATTGGGGAGTTTGCCGAA
GAATATTTGGACAAGGAGTCTGGAGAGACTGTAGATGAATTGGTGCAAAAGGTTGCTCGA
TACCTTCGTTCAAAGTCAAATACGGATACGTTCCCTGAAATGGAAGAGGAGGAAGCACCG
TTTACTAATCAGGAGAAGATTAGGTTCAATCGTCTCTTATCCCGAGCCTACGAAGAATCA
GGGATTGTCACATCAGCATTTGCCAAAACCTTTTACCTCGGTACACAAGTTCTTCCGGAA
CCATCCATGAAGGCCATTTGGGCAATTTACGTTTGGTGCCGTAGAACAGATGAAATCGTC
GATGCCCCTCGTCCTGCGGCTCACGATCCTAATGCAGAAATGTTGACAGATTTGTCGGAG
TGGGAGATTCGTCTCGAGAGACTGTTTGATCGAGGGGAGGTTGTGGATGTGCTGGACTTG
CCACTGTTGGACTGTAAGGTCAAGTATCCTACGTTGCCCATTACACCATTTTCAGATATG
ATACGTGGTATGTTGATGGACATCCCTGGGCTGGGACAAGAGAGATACGATACTTGGGAT
GAGTTGCACCTATATTGTTATCGCGTGGCCGGGACTGTGGGGTTGATGTCCATGCCAGTG
TTTGGATGTGCGGAAGGTTACACCGATGAAGTTGCAAA

Exon 2:
>Thaps3 chr_5:1330907-1330265
GGAGCCTGCTCTTTCGCTTGGTGTTGCATTCCAAATTACCAATATTCTCCGTGATGTAGG
AGAGGATGCTGAGAAACGCGAACGTGTATACCTTCCCCAACGAGACATGGAGAGGTTTGG
AGTTACTGAACGACAAATTTTTGACAAAGTAGTGGATGAGAACTACATTAATCTCATGAA
GTTTGAGATTGCTCGTGCCCGAATGTACTACGCTCGTGCTCTGAGAGGTGTGCCCATGCT
TCGCCCAGAGTCACGTCTGCCAGTGCAACTCTCGTTGGATGCCTACGGGAAGATATTGGA
CAAAATTGAGGAGAATGGATATGATTCGCTGACAAAACGAGCATATGTTGGTAAGTGGGA
GAAGTTGGCAGGAATCCCAGCGTCTTGGTACCGTACTCTTGATATTGCAAAGGCAATGCC
ACTCCCTGGGGATTGGGAGCGTCCGTCGTTAGAAGAATATGAAAAAAGTTTGGAACAAAT
GTTGGAGAATAAATGGGTAGAACAAAGGGGGAGCTCAGCGACAAGTGATTATTTGCAAAG
CCTCAACAAATAAGTGCAACTCGGCAAACCCTCAACGAAAGCATTGCTGCATCGCGAGGC
AGCGTTGAATAAAACAATAAATGTAGACAATGATGACGCAGTT

Translated:

MRISSIVAATLITCDVASAWSSSAFTSPLSIKTQHRSQASTKRRDGSILMSAVSQKESSSSSSGQG
RTTKEISNELALGVTLDGGPVIDFASVKDTTSRAEIALADSRKKYEASGATISPNPGGRL

MGINDEVVAEVGYEIGEFAEEYLDKESGETVDELVQKVARYLRSKSNTDTFPEMEEEEAP
FTNQEKIRFNRLLSRAYEESGIVTSAFAKTFYLGTQVLPEPSMKAIWAIYVWCRRTDEIV
DAPRPAAHDPNAEMLTDLSEWEIRLERLFDRGEVVDVLDLPLLDCKVKYPTLPITPFSDM
IRGMLMDIPGLGQERYDTWDELHLYCYRVAGTVGLMSMPVFGCAEGYTDEVAKEPALSLG
VAFQITNILRDVGEDAEKRERVYLPQRDMERFGVTERQIFDKVVDENYINLMKFEIARAR
MYYARALRGVPMLRPESRLPVQLSLDAYGKILDKIEENGYDSLTKRAYVGKWEKLAGIPA
SWYRTLDIAKAMPLPGDWERPSLEEYEKSLEQMLENKWVEQRGSSATSDYLQSLNK-

Targeting:

SignalP 3.0 - NN = Yes, Cleavage Site ASA-WS

SignalP - HMM = Yes, Cleavage Site ASA-WS, Signal Peptide Probability = 0.993

SignalP 4.1: Yes, Cleavage Site ASA-WS, D = 0.554 (D-cutoff = 0.450)

ChloroP: Yes, Score = 0.558

2)Thaps3_263269 (PSY2)

No introns.

>Thaps3 chr_7:1643508-1642761
AAAAAAAGTCGGATTGATACAACCAGAAGCGAAGTACACCGACCCAACAACAACCAACAC
CCCAACACCTCGGTCTGATCTGCTCTGCACACTCCAATGATGCTCCTACTATTATCAATA
TCGGCGTTATGCAACATCTCACTTGCTCGAGCGTTTGGGAACCTCTCACCAACGAACCGC
ATCAACCACCACCTCGACCTCTCCTCCTTTGTCCACCTATCTCGAACTCGTCTATATAGC
AGCCCACATTCAGCCAACCCAACAATATCCAATGAAATTATACAACTCATGTCAAAACAC
GATCCCATTCTATTATTTGCATCGCGACTGTTACCGTATCAAACTGCGGTCGACGCGTCG
GCTTTATACGCATGGTGTCGAAGATTGGATGAAATCACAGATGATCCATCAGCTAATGTG
CACTCTATACAACAACAACTGATTGATTGGGAGGATCGATTCAAGAAATTGTGCACTGGA
CAGCCTGTGGATGAAATGGACGATGCATTGTATCAATGCTTACAACGAAACTCAAACTCT
CTAAATGAGCGACCTTTTCAAGACATGATTGTGGGTATGAAGAGTGATGCTGTTCCAACT
ATTCGTACAATAAGCAGTATGGAGGAGTTGGAAGAATATGCCTATCAAGTCGCTGGAACT
GTGGGATTAATGCTGCTTCCGGAATCGTGTGGAGAAGGCACGTCAGCCTGCCATTGCTCT
TGGAAAGGCCATTCAACTGATTAACATA

Translated:

MMLLLLSISALCNISLARAFGNLSPTNR
INHHLDLSSFVHLSRTRLYSSPHSANPTISNEIIQL MSKHDPILLFASRLLPYQTAVDAS
ALYAWCRRLDEITDDPSANVHSIQQQLIDWEDRFKKLCTGQPVDEMDDALYQCLQRNSNS
LNERPFQDMIVGMKSDAVPTIRTISSMEELEEYAYQVAGTVGLMLLPESCGEGTSACHCS
WKGHSTD-

Targeting:

SignalP 3.0 - NN: Yes, Cleavage Site ARA-FG

SignalP 3.0 – HMM: Yes, Cleavage Site ARA-FG, Signal Peptide Probability = 1.000

SignalP 4.1: Yes, Cleavage Site ARA-FG, D = 0.826 (D-cutoff = 0.450)

ChloroP: No, Score = 0.489

If the first methionine is omitted, the ChloroP score becomes 0.499.

3)Thaps3_23291 (PDS1)

No introns.

>Thaps3 chr_6:1795938-1794225
AACGCAAAGCGTTCCTCTTCGGCGTCAGCCTCCTCCTTTCTTTCACTGTCAGTCAGTTGT
GTTGAACACTTCTGCTTCTTCATTCATCTCCTCATTACTACCACTGCTATTGCAACTGGC
TGCCGTATCTGACTCTCCTTCAATTCGTCACTCCCATAGCGCCACGTTCACCATCATTTA
TCACCATGATCATTACAAATTTCATCCTCTCCACCGTCCTAGCGACATCAATGGCCTTTC
AACCACACACACCCATCCTCTCCAAACCATCCTTCTCCAACCGTGTCCATCGCTCCCCCA
AAATCGGCTCTTCCAACCTCGTTATGAAGGACTTTCCGAAACCAAATGTCGAAGATACAG
ACAACTATCGCTACGCAGAGGCCATGTCCACTAGCTTCAAGACGTCTCTCCGAGTGACGA
ATGATTCACAGAAGAAGAAGGTGGCTATCATTGGAGGAGGATTATCAGGTCTGTCTTGTG
CCAAGTACCTCTCCGATGCCGGGCATGAACCCACCGTATACGAAGCACGTGATGTACTCG
GAGGAAAGGTGTCAGCGTGGCAAGATGAAGATGGAGACTGGATCGAAACAGGTCTTCACA
TCTTCTTCGGAGCATACCCCAACGTTATGAACATGTTCGCTGAGCTTGGCATCCACGATA
GGCTTCAGTGGAAGATTCACCAAATGATTTTCGCAATGCAGGAACTTCCCGGAGAGTTCA
CTACCTTTGATTTCATCCCTGGTATTCCAGCTCCGTTCAACTTTGGATTGGCCATTCTTA
TGAATCAAAAGATGTTGACGTTGGGTGAAAAAATTCAGACCGCTCCTCCTCTTCTTCCTA
TGCTTATTGAGGGACAGTCATTCATTGATGCTCAGGATGAGTTGAGTGTGACGCAGTTCA
TGAGGAAGTACGGTATGCCTGAGAGAATCAACGAGGAGGTGTTTATTGCGATGGCCAAGG
CGTTGGACTTTATTGATCCTGATAAGTTGAGTATGACTGTGGTGCTTACGGCTATGAACA
GGTTCTTGAATGAGAGTAATGGACTTCAGATGGCATTCTTGGATGGAAATCAGCCTGATA
GGTGGTGCACTCCCACGAAGGAGTATGTGGAAGCACGCGGAGGAAAGGTCAAATTGAACT
CTCCCATTAAGGAGATTGTGACCAACGACGATGGAACTATCAATCACCTTCTCCTTCGAT
CTGGCGAGAAGATTGTGGCCGATGAATACGTCTCTGCCATGCCCGTGGACATCGTCAAAC
GTATGCTTCCCACAACGTGGCAGACTATGCCCTACTTCCGTCAGCTTGACGAACTTGAGG
GCATCCCTGTTATCAACTTGCACATGTGGTTCGATCGTAAGTTGAAAGCAGTCGACCATC
TTTGCTTCAGTCGCTCCCCACTCCTTTCCGTCTACGCCGACATGTCCGTCACATGCAAGG
AGTACGAAGATCCCAACAAGTCCATGTTGGAATTGGTCTTTGCTCCCTGCTCTCCTATTG
CCGGAGGAAATGTCAACTGGATTGGAAAGTCAGATGAGGAAATCATTGATGCTACCATGG
GTGAGCTTGCTCGCCTTTTCCCTACCGAGATTGCGAATGATGATAAGTGGCCTGCTACGA

AGATGCAGGGACCTAATGGACAGGCAAAGCTTGAGAAGTATGCTGTTGTGAAGGTGCCAA
GGAGTGTGTATGCTGCCATTCCTGGTGAGTGAAA

Translated:

MIITNFILSTVLATSMAFQPHTPILSKPSFSNRVHRSPKIGSSNLVMKDFPKPNVEDTD
NYRYAEAMSTSFKTSLRVTNDSQKKKVAIIGGGLSGLSCAKYLSDAGHEPTVYEARDVLG
GKVSAWQDEDGDWIETGLHIFFGAYPNVMNMFAELGIHDRLQWKIHQMIFAMQELPGEFT
TFDFIPGIPAPFNFGLAILMNQKMLTLGEKIQTAPPLLPMLIEGQSFIDAQDELSVTQFM
RKYGMPERINEEVFIAMAKALDFIDPDKLSMTVVLTAMNRFLNESNGLQMAFLDGNQPDR
WCTPTKEYVEARGGKVKLNSPIKEIVTNDDGTINHLLLRSGEKIVADEYVSAMPVDIVKR
MLPTTWQTMPYFRQLDELEGIPVINLHMWFDRKLKAVDHLCFSRSPLLSVYADMSVTCKE
YEDPNKSMLELVFAPCSPIAGGNVNWIGKSDEEIIDATMGELARLFPTEIANDDKWPATK
MQGPNGQAKLEKYAVVKVPRSVYAAIPGE-

Targeting:

SignalP 3.0 - NN: Yes, Cleavage Site SMA-FQ

SignalP 3.0 - HMM: Yes, Cleavage Site SMA-FQ, Signal Peptide Probability = 0.864

SignalP 4.1: Yes, Cleavage Site SMA-FQ, D = 0.603 (D-cutoff = 0.450)

ChloroP: Yes, Score = 0.550

4)Thaps3_1383 (PDS2)

Exon 1:
>Thaps3 chr_1:1373161-1372917
CGGGGAACAGACCTCAGACGTCCTTAAGTTGGATCCTTCTCTCATCCCTACTAACCAATC
ATGAAGTTTCTTCTACCGCTGCTACCAGCAGTCGCTGGGGCCTTCTCCATAACGCACCTC
TCACAACACCCATCGTTACGCATGCATCAATCTTTGTCCACATCGCTGTACTCTTCCTCC
TCCTCCACATCGCAACGTCCAAGACGTCCAACTCCTGATCGTATTCGCAATACACAAAAC
TTTAA

Exon 2:
>Thaps3 chr_1:1372807-1372145
AGAAGCGAAAGAGTTGAGTCAGAAATTCATAACAGACTTTCAACAACTACAAAAGGTTGG
TAGTGGCGAACCGAAGCGAGTTGCCATATTCGGAGGTGGTCTTTCTGGTCTTTCGTGTGC
AAAGTATCTCTCAGATGCTGGTCACATCCCAACCCTGTACGAGGCTCGTGGTGTTCTCGG
GGGCAAGGTCTCGGCATGGCAAGATGAAGACGGAGATACAGTCGAGACCGGACTACACAT
TTTCTTTGGTGCTTATCCAAACATCCACAATCTCTTTGATGGGCTGAAAATACAAGACAG
ATTGCAATGGGCTCCTCACAGAATGACATTTGCCATGCAAGAGCTTCCCGGGCAATTTAC
CACCTTTGAGTTCCCTGCTGGCGTTCCTGCTCCATTGAATATGGCTGCTGCGATTCTGGG

GAATACTGAAATGCTTACGTTGGAAGAAAAGATTAAAATGGTTCCAGGGCTATTACCAAT
GCTATTGGAAGGGCAATCTTTCATTGACGAGCAAGATGAGCTTTCTGTTTTGCAATTCAT
GCGGAAGTATGGTATGCCCGAACGTATCAATGAAGAGATATTTATTGCAATGGGAAAAGC
ACTGGACTTCATCGACCCTGATCTACTGTCCATGACGGTTGTTCTCACGGCAATGAATCG
TTT

Exon 3:
>Thaps3 chr_1:1372066-1371065
CATCAACGAAGCAGACGGAAGCCAAACAGCTTTCCTTGATGGGAATCCCCCAGAGCGACT
ATGTCAACCTATGAAAGAGTCCATCGAAAAGAAGGGAGGAGAAGTAGTTTGCAACAGTCC
TGTAGTTGAGATTCAACTGAACGAAGAGAGTAACGTCAAGTCTCTCAAACTTGCAAATGG
AACTGAAATCACAGCAGATTATTACGTGTCGGCAGTGCCTGTGGATGTCTTCAAACGTCT
CGTGCCCACGCAGTGGTCAACAATGCCTTACTTTCGTCAACTTGATGAACTTGAAGGAAT
ACCTGTCATCAACATTCAGATTTGGTTTGACCGAAAGCTCAACTCGGTGGATGGATTGTG
CTTTAGTCGGTCTCCACTGTTGAGTGTCTATGCGGATATGTCAACGTGTTGCGAAGAATA
TGCAAGTAACGATAAATCCATGTTGGAGTTGGTGTTTGCACCGTGTTCTCCTGAGGCGGG
ATCTCCATTGAATTGGATTGCGAAGCCAGACTCTGATATCATTGACGCAACAATGAAGGA
GTTGGAGCGCCTCTTTCCCTTGGAGATCGGTCCCGATGCTCCCGAGGAGAAACGCGCCAA
TGTTGTAAAGTCTACGGTGGTCCGTGTACCTCGAAGTGTGTATGCGGCTGTTCCTGGCAG
AAACAAATATAGACCTAGTCAGGAATCACCAATCGAAAACTTCATTATGGCTGGAGATTA
TGCAACACAGAAGTACCTTGGTAGCATGGAGGGGGCTGTACTCTCAGGGAAACTTGCAGC
TGAGGTCATTTGCGACAAGTTCATGGGCAGAGCGGAGAGGAAAGGGGTCAAAGAGGTACA
CTCATCGGTGCTCACGAAGCAAATCGAAGAGAGAACCCCAGCGGGTATTGCAATGGAAAA
AGGCAGAGTGTCGCCAACATCGTATGGAGGTGGTCAACAAGGTGGCTTTGAAAATCCCTA
AAAGTTTAGACTAATGCACTGTACAATCTGCTCAGGATGTTT

Translated:

MKFLLPLLPAVAGAFSITHLSQHPSLRMHQSLSTSLYSSS
SSTSQRPRRPTPDRIRNTQNFKEAKELSQKFITDFQQLQKVGSGEPKRVAIFGGGLSGLS
CAKYLSDAGHIPTLYEARGVLGGKVSAWQDEDGDTVETGLHIFFGAYPNIHNLFDGLKIQ
DRLQWAPHRMTFAMQELPGQFTTFEFPAGVPAPLNMAAAILGNTEMLTLEEKIKMVPGLL
PMLLEGQSFIDEQDELSVLQFMRKYGMPERINEEIFIAMGKALDFIDPDLLSMTVVLTAM
NRFINEADGSQTAFLDGNPPERLCQPMKESIEKKGGEVVCNSPVVEIQLNEESNVKSLKL
ANGTEITADYYVSAVPVDVFKRLVPTQWSTMPYFRQLDELEGIPVINIQIWFDRKLNSVD
GLCFSRSPLLSVYADMSTCCEEYASNDKSMLELVFAPCSPEAGSPLNWIAKPDSDIIDAT
MKELERLFPLEIGPDAPEEKRANVVKSTVVRVPRSVYAAVPGRNKYRPSQESPIENFIMA
GDYATQKYLGSMEGAVLSGKLAAEVICDKFMGRAERKGVKEVHSSVLTKQIEERTPAGIA
MEKGRVSPTSYGGGQQGGFENP-

Targeting:

SignalP 3.0 - NN: Yes, Cleavage Site AGA-FS

SignalP 3.0 - HMM: Yes, Cleavage Site AFS-IT, Signal Peptide Probability = 0.994

SignalP 4.1: Yes, Cleavage Site AFS-IT, D = 0.465

ChloroP: Yes, Score = 0.569

5)Thaps3_24832 (ZDS)

No introns.

>Thaps3 chr_14:437250-439255
CTCAACCAACGAAGAAGCCTTTTTTTCACTTGTCGTCAGCTCTCTCATTCTCTGCGACCC
AACCCATAGAGACTGCCGAAAACAATTCACATCCATCGCACCTCACATAGCCACTCCGAA
ACACCTGCCATGCGTCTCTCCACAGCCTTCCTCGTCGGATGTGTCCTCCCAGCAACACAC
TCGTTCCACCTTCCCTCCGCCAGCACCTCTCTCCGTCGTCCCGTCACCGTCGCCACCACA
TCATCCCTCTCCATGTCCGCCGCCACCTCTGAAGGAGAATTCACCTCCGAATCCGCCAAA
CAACAAATCGGAAACGACTCCTTCCTCAACGAAAACCTCATGGCGCGTGCCCAAAACGGA
CCTGGCAAAGTCAACGACGAAAAGCTCAAGATTGGAGTAGTCGGCGCTGGTTTGGCGGGT
ATGGTCGCCGCGATGGACTTGGCCGATGCAGGTCATGACGTGGAGATGTTTGAGTTGAGG
CCTTTTGTGGGGGGGAAGGTTTCGTCGTGGAAGGATAAGGAGGGGAATCATATTGAGATG
GGGTTGCATGTGTTCTTTGGGTGTTATTATAACTTGTTTGGGATTATGAAGAGAACGGGA
TCGTTTGATACCGAGTTGAGGATTAAAGAGCATATTCACACTTTTGTGAATGAGGGTGGG
ATTTTGGGTGCGTTGGACTTCAAATTTCCTATTGGGGCTCCTATTTCGGGACTTCAAGCC
TTTGCTCGTACGGAGCAGTTGGGATGGGATGACAAGTTCCACAATGCACTGAGGTTGGGT
ACTTCTCCTATTGTGAGGGCGCTGTTTGACTTTGATGGTGGTATGGATATGGTGCGAGAC
TTGGATGATATCACGTTTACTGAGTGGTTTACTCAGTTGGGAGGATCACGAGGAAGTTTG
GATAGGATGTGGGATCCCATAGCGTATGCGCTTGGTTTCATTGATTGCGATCACATCTCG
GCAAGGTGCATGCTGACAATCTTTATGCTTTTCGCCATCAGAACTGAAGCGAGTGTGTTG
AGAATGCTGGAAGGAAGTCCTCAGACGTGCTTGCACGATCCTATTCTCAAGTATTTGGGT
GATCGTGGGGTCAAGATCAATACCTCCATGGGATGCAGAGAGATCGTGCACGATGTGGAT
GAGAATGGCAAGCCTATTAGAGTGACTGGAATCAAGGTTGGACCCAAGGAGGAGTTGAAG
GAGTTTGATGCCGTGGTTTGTGCATTGGATGTTCCTGGAATCAAAAAGGTCTTGCCTCAA
TCATTCAGGGATCACTACCCAATGTTTGACAACATTTACAACCTCGACACTGTTCCTATT
GCTACCGTTCAAGTCCGATTCGATGGATGGGTGACTGAAATGAATGACGACGTTCGCATG
ATGGATATTTCTGGAGATCAATCCGACGGACGTGGCGGTGGAATTGACAACCTTCTCTAC
TCTGCCGATGCTGAGTTCTCATGCTTCGCTGATCTCGCCATCACTTCTCCCGGAGAGTAC
TACAAAGAAGGGGAGGGAAGTCTCATCCAAGCAGTGTTTGACGAACGTGCCTTTGATCGT
TCCAATGACCAAATTGTCCAAGATTGCATCAGTCAATTGAACTCACTGTTCCCTTCCAGT
AAGAAGCTCAATTGCACCTGGTCAAGTGTGGTCAAGTTGGGACAATCACTGTACAGAGAG
AAGCCAGGACAGGACAAGTTCCGTCCGAAGCAGGCTACTCCCATCTCCAATTTTTTCTTG
GCTGGAAGTTACACTTACCAAGATTACTTGGATTCCATGGAGGGTGCTACGCGTAGTGGT

TTGATGGTGGCGGATGAGATTATTGCAAGAGCTGACGGACCTAATGGATTGAAGGCACAG
ACTGCCAAAGCGATGGCCAATGGGAGTGGGAAGAAGGCTATTAGTGAGGAAAAGGTTCCA
TTCTTTGCATCAGCGTCGGCATAATGGGGGTACAGCAGGCGGAGGATAATGTTCTTAATG
ATATTTAGTTGAAATACTAAGGCAGG

Translated:

MRLSTAFLVGCVLPATH
SFHLPSASTSLRRPVTVATTSSLSMSAATSEGEFTSESAKQQIGNDSFLNENLMARAQNG
PGKVNDEKLKIGVVGAGLAGMVAAMDLADAGHDVEMFELRPFVGGKVSSWKDKEGNHIEM
GLHVFFGCYYNLFGIMKRTGSFDTELRIKEHIHTFVNEGGILGALDFKFPIGAPISGLQA
FARTEQLGWDDKFHNALRLGTSPIVRALFDFDGGMDMVRDLDDITFTEWFTQLGGSRGSL
DRMWDPIAYALGFIDCDHISARCMLTIFMLFAIRTEASVLRMLEGSPQTCLHDPILKYLG
DRGVKINTSMGCREIVHDVDENGKPIRVTGIKVGPKEELKEFDAVVCALDVPGIKKVLPQ
SFRDHYPMFDNIYNLDTVPIATVQVRFDGWVTEMNDDVRMMDISGDQSDGRGGGIDNLLY
SADAEFSCFADLAITSPGEYYKEGEGSLIQAVFDERAFDRSNDQIVQDCISQLNSLFPSS
KKLNCTWSSVVKLGQSLYREKPGQDKFRPKQATPISNFFLAGSYTYQDYLDSMEGATRSG
LMVADEIIARADGPNGLKAQTAKAMANGSGKKAISEEKVPFFASASA-

Targeting:

SignalP 3.0 - NN: Yes, Cleavage Site THS-FH

SignalP 3.0 - HMM: Yes, Cleavage Site THS-FH, Signal Peptide Probability = 0.999

SignalP 4.1: Yes, Cleavage Site THS-FH, D = 0.543 (D-cutoff = 0.450)

ChloroP: Yes, Score = 0.572

6) Thaps3_270357 (LCYB)

No introns.

>Thaps3 chr_2:1232356-1230385
CGGGTTGACACGTCATCAACCTTTCTATCACTCGTCCTCGCTCGGCATTCTTCCTCATGG
TGGCCGTCAATCCTCTCCACGCTGTGGCATTAGCAGTGCTTCCATCGGCGGCATATGCCT
TTGTATCTCCTACACCAGTCGTCCTCTCCCCGTCGTCGGCTAGAGTGACACCCTTGTCTG
GTGCACCTTACTGCACCTCATATCGCGTTCGTCATTCCCACTCCTTCCCGACAACGTCCA
TCCAATCTCGTACCACCTCCTCTACAATCTTTGCATCCGCGTCCAACCCAACTCTCACTC
CAACTTCTCAAGATACATGCGATGTTCTCGTCCTCGGAAGCGGACCTGCAGCACGATCCA
TCGCCACCCTCCTCTCCTCCAAAGCAAACGACAAAGCATACGACGTACTTCTAGCAGACT
CAAATTACGATCGCAGATGGGCTCCTAACTACGGTGTTTGGCAAGATGAATGGCAATCTA
TTTGCAAACTATATGAATCATTCAATCAGCCCATTGGTAATGAAGTTATTGATCGGTTGT
GGATGTCGACGGATTGTTTCTTCGGTGGCTCTTTTGACATCGCTGCGGAGCAACGGATGA
GATTGGATAGGCCATATTGCAGGATTGAGAGGGATGTGCTGAGAAGAGTGTTGAGTCCAG
CGAGTGAAGTGAAGACACAGAACGATGGAGAGGGGACAGCAAACTATCGTGTTACACGTG

CCAATCACATGTCAAAGTGCACGTCTGTGAACATTTATTCTCCCTCTGGATCAATGGTAC
ACGATGAATCTGGAACCAGTGTTTTGATTCAATCAAAGGATGGTGATATTTCTCAAGTAC
GTGCAAAGCTAGTTGTGGATTGTACGGGACATGAATCCAAAATCGTGTTGAAGGATGATC
GTATGAAATCCATTCCTCCAGGCTTTCAGATTGCCTACGGTATTCTAGCAGAAGTTGACG
AGACGAGCATACCAAACAATGATTTCTGCGGTCCGTACTTCAAAGAGGCAATGACTTTGT
TTGATTACAGAACCGATCACTTTCCGGAAGGTTCTAATGAGTTGGCTAAAGCAGAAAAGG
CACCGACGTTCATGTACGCCATGCCACTCAATGGGAATCGTATCTTTTTCGAAGAGACAT
CGTTAGTGGCTAGACCAGCATTATCATTCCAAGAATGCAAAGATCGATGCATGACACGTC
TCGAGCACTTGGGGATTACTGTTACAAAGGTTGAGGAAGAGGAGTTTTGCTACATTCCAA
TGGGAGGACCGTTGCCAGCCAAAGACCAAAGAGTTATTGGATTTGGAGGTGCGGCTGCGA
TGGTTCATCCTAGTACTGGGTACCACCTGTGTCGAGCAATGATGGGAGCAGGAGAAGTAG
CCAAAGTGATTCGTGAAGAGTTGGAAGAGAAGAACTGGAACCCAGATAGGGCAGCTGCAC
GTGCTTACAATGCAATTTGGTCACCAACGACCATTGCCCAACGAAACTTTGCAGTCTTTG
GAGGTGAATTCTTGATGAAGCAAAACGTAGTAGGGCTGCGAGGCTTCTTTGATGGCTTCT
TCAAGCTTCCTCTTGGCTTGTGGGGAGGGTTCCTCGCTGGATGGCCTGGACTGCCAAACA
ATGAGAATCACGAGACGTGGTGGGCACGGTTAGTGTTTGGGTTGACGTTTGTTTCAAAGC
TACCCGTGTCTGTGGCGGTAGATATGCTTGGATCGATAGCAACATACTCCATTTCAGAGG
GAGTACCACTGCCTCAGTCGGTGACACCGTTATTGGGATTGCCTGATGGATATGAGTACA
GGGAAAAGAAGAGTTCTATAGGTGATGTTGCAGCGAAGCATGAGGCAAGGAGAATGATTA
TGGAGTCGACTGTTGAGGAAGTAGTTCCTGTAGACTTTGAAGAAAAGACAGCGGCGTGAT
TGGATTGATAACTCTTATTTGGAATGTAGTGATAGATACACAACTACAGATT

Translated:

MVAVNPLHAVALAVLPSAAYAFVSPTPVVLSPSSARVTPLSG
APYCTSYRVRHSHSFPTTSIQSRTTSSTIFASASNPTLTPTSQDTCDVLVLGSGPAARSI
ATLLSSKANDKAYDVLLADSNYDRRWAPNYGVWQDEWQSICKLYESFNQPIGNEVIDRLW
MSTDCFFGGSFDIAAEQRMRLDRPYCRIERDVLRRVLSPASEVKTQNDGEGTANYRVTRA
NHMSKCTSVNIYSPSGSMVHDESGTSVLIQSKDGDISQVRAKLVVDCTGHESKIVLKDDR
MKSIPPGFQIAYGILAEVDETSIPNNDFCGPYFKEAMTLFDYRTDHFPEGSNELAKAEKA
PTFMYAMPLNGNRIFFEETSLVARPALSFQECKDRCMTRLEHLGITVTKVEEEEFCYIPM
GGPLPAKDQRVIGFGGAAAMVHPSTGYHLCRAMMGAGEVAKVIREELEEKNWNPDRAAAR
AYNAIWSPTTIAQRNFAVFGGEFLMKQNVVGLRGFFDGFFKLPLGLWGGFLAGWPGLPNN
ENHETWWARLVFGLTFVSKLPVSVAVDMLGSIATYSISEGVPLPQSVTPLLGLPDGYEYR
EKKSSIGDVAAKHEARRMIMESTVEEVVPVDFEEKTAA-

Targeting:

SignalP 3.0 - NN: Yes, Cleavage Site AYA-FV

SignalP 3.0 - HMM: Yes, Cleavage Site AYA-FV, Signal Peptide Probability = 0.996

SignalP 4.1: Yes, Cleavage Site AYA-FV, D = 0.670 (D-cutoff = 0.450)

ChloroP: Yes, Score = 0.570

7) Thaps3_263437 (BCH)

No introns.

>Thaps3 chr_8:927506-925618
CGTCATCAGTCACCTTTCTTTCTCACCTCCATCATCCACCCTGCCATCCATCATCCACCC
TCTACCACCCTATCATCTACATACATTCACCCATGCATCAGCTATCCAGACTGCTTTCCA
AACTCTTCCCAGTCATCGCTACCTACATCATTGCGAAGTATGCTCTCCCTCACATCAATA
CGATCCTTCACTGTTCAAATACCTTGAGCAACTTTGTGAGAATAACATACACCGTCCTCT
TTGCAATAGCGATGGAGTACATTTCAAGATATAGTCACTGCTACTTGTGGCATGGAAAGT
TCCTTTGGTGGATTAATGGCTCACATCATCATCAGTATCCAGCTGTTGGAAGCACACCTG
TGTATGGACACAACAATCCGTACGTCTCTCCAGCCATTGAGTTGAACGATGCATTTGCCG
TTTTCTTTGCAACGATTGCCACATTGGCAATGTGGATAGGGTCAGAGCCTCCATCAACCT
TGACAAAAGATTGTTCTATTGGTATTGGATTGGGAGTGACTCTCTATGGCCTTTCATACT
TTGTCGGTCACGACATTGTTGCTCACGAACGTTTGGGCAAGGGAGTAGCAAACGCTTTGA
GACGAGCATTTCCGTATATGGAGCAATGTGCTTCTGTTCATATCCGGTATCATCACAAGT
TGACGAAACGTAGTAATGATTCTGATCCATATGGAGCGCCTTACGGCTTTTGGTTGGGTC
CGTCGGAAGTAGAGTGTTTGAACAGAGGTCAATGGTATGCACCGATGCCTATGTCGTTGA
AAGCTATTTCTTGGATAGCAACATTGATCTTCTTCGCATCGACAATTCACAGCTCGTTGT
CTCCGGCAGCTCAAGCAATTGTGCTCCTCGGATGTGTAGGATGGTGTGGCTCCGGCACAT
TGTCAGACAATACAACAACATCTCGTCGTATCGGCAAGCTGCTCTCATTGAACTGGCAAT
CAACTCCATCGAGACTGATGCCTCACGGACTTTCTGGGCTCATTTCAGTTGGCATTGGTT
CATACCTCATCTTTGGCCACAGTCTCGTAGGCGATCTGAAACCATACACAATGCAACAAC
CACCATATCTCATCATTCTCTATGCAACAGCCACATCATGGAATGCTCTCGGAGGATATA
TGATTGTCAATACTGCTCCTCCAAATACAAGAATGCTGTTCAGGAGATGTGCAATACTTC
AGGTTTGCTTGTCGTACTTCATTGTTAGGTTCCTTCCGCACTCGTCGGTGCTGCTGATAC
GTCTCGAGAGCAACACAATTACAACATCGCTGAGATGTCTTGACTTAATCGTTACGATAT
CAGCAGTGGTGTGTACTCTATCGTTCTTTGATGCAGTCGTTGACATGTCAAAGCAGAGCG
TTGTCCTCGGACAATCAATCGCATTCGGTATCATCGGGATATTGCTCCTTTCCGTGTATC
CAATCCAACTGAGTTTGCAAGGAGAAGAGTGGTGGAGTTGTATACAAAACAGATATCCAA
TGCAAGCCAGTGGAATGATTGCTTATATTTACGTTCCAGCTACAGTGACGTTTAGTTTGT
TTCTCTTCGGAGCAACGCTGTATCAAAGAAAGATAATGTCTGCATCAGAGTATGGGATTA
TATCATTGATGGTGATACTTGTATGCCTATTGGCAACTGTGCTCAGCCAAGAGATACATA
TCCCAGACGTTTCGACGCAGAGAATCTATTTACCATGCGAAGATCCAGCAATGGATTCGT
TGGAAGAGAAAGTACTTGAAGCGTTGGACTTTTCTCGTTATGCACGTTCCATCTTGACCA
CAGTCTTGGGTATTAAATTTGAGTCACCCGCGTAGCATGATGGGAATCTCCAATTTAACT
AATTGCACATTAATCATTGTTGTATTATC

Translated:

MHQLSRLLSKLFPVIATYIIAKYALPHINT
ILHCSNTLSNFVRITYTVLFAIAMEYISRYSHCYLWHGKFLWWINGSHHHQYPAVGSTPV
YGHNNPYVSPAIELNDAFAVFFATIATLAMWIGSEPPSTLTKDCSIGIGLGVTLYGLSYF
VGHDIVAHERLGKGVANALRRAFPYMEQCASVHIRYHHKLTKRSNDSDPYGAPYGFWLGP
SEVECLNRGQWYAPMPMSLKAISWIATLIFFASTIHSSLSPAAQAIVLLGCVGWCGSGTL

SDNTTTSRRIGKLLSLNWQSTPSRLMPHGLSGLISVGIGSYLIFGHSLVGDLKPYTMQQP
PYLIILYATATSWNALGGYMIVNTAPPNTRMLFRRCAILQVCLSYFIVRFLPHSSVLLIR
LESNTITTSLRCLDLIVTISAVVCTLSFFDAVVDMSKQSVVLGQSIAFGIIGILLLSVYP
IQLSLQGEEWWSCIQNRYPMQASGMIAYIYVPATVTFSLFLFGATLYQRKIMSASEYGII
SLMVILVCLLATVLSQEIHIPDVSTQRIYLPCEDPAMDSLEEKVLEALDFSRYARSILTT
VLGIKFESPA-

The underlined portion aligns with BCHs from plants and green algae **(Appendix D)**.

Targeting:

SignalP 3.0 - NN: Yes, Cleavage Site VIA-TY

SignalP 3.0 - HMM: No, Signal Peptide Probability = 0.299

SignalP 4.1: No, D = 0.285 (D-cutoff = 0.500)

ChloroP: No, Score = 0.458

8) Thaps3_9541 (LTL1)

No introns.

>Thaps3 chr_13:695539-693528
CTCCAACTTCGCATCGGTCAACACAACACTGCAAGAGCGAAGGAAAAGCCAAACCCTAAC
ATGAAGTTCACAACAGCTCTCGCCGTCCTCTGCTGGACGTCGGTGACAAATGCCTTCGTT
CCATCGTCCTTCACCTCTCCAGCGTTGAAGAATGAGCAGCAGCAAGTACGTGCATCATCT
CCACTGTACGCACTTGATACCAAAGAAAAGGAAGAAACCACCACGGCCACCTCTGCTTCC
TCTACCGACACATCTTCCACACCAGCAGCAGCAGCCACTGAAGAATCCGAAGGACTCCCA
TGGTGGTGGGAATACATCTGGAAGCTCCCCGTAATGCAACCAGCCGAACCAGGCACCGAC
ATCATCTTCGCCGACAGTGCTCGCGTCCTTCGCACGAACATTGAACAAATCTACGGAGGA
TTCCCCTCCCTCGATCAATGTCCATTGGCCGAAGGAGAAATTACCGATATTGCCGATGGA
ACAATGTTCATCGGTTTGCAGAGGTATCAACAACAGTATGGAAGTCCGTACAAACTGTGC
TTTGGTCCAAAGAGCTTCTTGGTGATTTCGGATCCAGTTCAAGCCAAACACGTTCTACGC
GATGCCAACACTCTCTACGATAAAGGAATCTTGGCTGAGATTCTTAAACCGATCATGGGG
AAGGGGTTGATTCCAGCAGACCCAGAAACGTGGTCGGTACGACGTAGGGCGATTGTTCCT
GCCTTTCACAAGGCGTGGTTGAATCATATGGTGGGATTGTTTGGTTATTGCAATGAAGGA
TTGATTGCTTCGTTGGAGGAGGCGGCGAAGAAAAATGATGCTCCTAACGGACAACAGGGT
GGAAAGATTGAGATGGAAGAAAAGTTTTGCAGTGTGGCACTGGACATCATTGGGTTGTCG
GTGTTCAACTATGAATTTGGATCAGTGAGCGAAGAATCACCAGTGATCAAGGCAGTGTAC
TCTGCATTGGTGGAGGCAGAACATCGTAGCATGACTCCCGCTCCTTACTGGGATTTGCCA
TTTGCCAACGAGGTGGTACCACGTCTACGCAAGTTCAATAGCGATCTCAAGGTCTTGGAT
GATGTGTTGACTGATTTGATTGATCGTGCCAAGAACTCACGTCAAGTGGAGGACATTGAA
GAGTTGGAGAAGCGTGATTATGCCAATGTGAAGGATCCATCGTTGTTGCGATTCTTGGTG
GATATGAGAGGTGCTGATATTGATAATAAGCAATTGAGGGATGATTTGATGACGATGCTT

ATTGCAGGGCACGAGACGACTGCGGCTGTGTTGACTTGGGCGTTGTTTGAGCTAACAAAG
CATCCGGAACAGATGGCCAAGGTTCGTGCCGAGATTGATTCTGTGTTGGGAGATCGTACA
CCGACATATGACGATATCAAAGAGATGCAGTATTTGAGGTTGGTGGTTGCAGAGACTTTG
AGGTTGTATCCTGAGCCTCCGCTGTTGATTCGTCGTTGTAGAACCGAAAACAAGTTGCCC
AAGGGAGGTGGAAGGGAGGCTACTGTGATTCGTGGTATGGATATCTTTCTATCATTGTAC
AATCTTCACCATGATGAGAGGTTCTGGCCGGAGCCTAATGAGTTCAAGCCCGAGAGATGG
GAGAGCAAGTATATCAATCCTGAGGTACCAGAATGGGCTGGATATGATCCTGCAAAGTGG
ATAAATACCAACTTGTACCCCAACGAGGTCGCGTCCGACTTTGCCTACTTGCCATTCGGA
GGAGGAGCACGAAAGTGTGTCGGAGACGAGTTCGCCACGTTGGAGGCAACTGTGACTTTG
GCAATGTTACTCCGTCGCTTTGAGTTCGAGTTTGACTCTGCCAAGCTTGCTGCTTCGAAG
ATTGACATTATGGATCATCCAGAGGATTTGGAGCATGCGGTTGGTATGAGGACCGGAGCT
ACTATTCATACTAGGAAGGGATTGCATATGGTGATTAGGAAGCGTGAGTTGTAGTTCATT
AGAGCCTAGTCTAAGCAACAACTATGGTCATT

Translated:

MKFTTALAVLCWTSVTNAFVPSSFTSPALKNEQQQVRASS
PLYALDTKEKEETTTATSASSTDTSSTPAAAATEESEGLPWWWEYIWKLPVMQPAEPGTD
IIFADSARVLRTNIEQIYGGFPSLDQCPLAEGEITDIADGTMFIGLQRYQQQYGSPYKLC
FGPKSFLVISDPVQAKHVLRDANTLYDKGILAEILKPIMGKGLIPADPETWSVRRRAIVP
AFHKAWLNHMVGLFGYCNEGLIASLEEAAKKNDAPNGQQGGKIEMEEKFCSVALDIIGLS
VFNYEFGSVSEESPVIKAVYSALVEAEHRSMTPAPYWDLPFANEVVPRLRKFNSDLKVLD
DVLTDLIDRAKNSRQVEDIEELEKRDYANVKDPSLLRFLVDMRGADIDNKQLRDDLMTML
IAGHETTAAVLTWALFELTKHPEQMAKVRAEIDSVLGDRTPTYDDIKEMQYLRLVVAETL
RLYPEPPLLIRRCRTENKLPKGGGREATVIRGMDIFLSLYNLHHDERFWPEPNEFKPERW
ESKYINPEVPEWAGYDPAKWINTNLYPNEVASDFAYLPFGGGARKCVGDEFATLEATVTL
AMLLRRFEFEFDSAKLAASKIDIMDHPEDLEHAVGMRTGATIHTRKGLHMVIRKREL-

Targeting:

SignalP 3.0 - NN: Yes, Cleavage Site TNA-FV

SignalP 3.0 - HMM: Yes, Cleavage Site TNA-FV, Signal Peptide Probability = 0.999

SignalP 4.1: Yes, Cleavage Site TNA-FV, D = 0.797 (D-cutoff = 0.450)

ChloroP: Yes, Score = 0.548

9) Thaps3 270336 (LTL2)

Exon 1:
>Thaps3 chr_9:683521-684549
CGCCCTCCCCTCAACCACTTCCACGCTTTTTGGCTCTGGTTTAGCTGCTGATAAAAGTTG
AGAGTAGTCCAACAATACAGCAATACGACTCTCTTCGTTGCGTAGTAGTAGACTGTGTAT
CATCGGTGGAGGAAAGCTGAGTTGATTGGATAGAATTGGGCTGATACAACGGATAGAAAG
AACATCAAGATGTGCACCAAACTATCCAGCCGTCGGACGCTGTTGGCGTTGTACTTCGCC

TTTACTGGATGCACTGCCTTTCAACTACCATCGGCCACACCATCACGTGCGTCAATAACA
AAAGCATACAGTACCCACCTTGATAAAGAGATCAAAAGCAAAACGCCTCTCGTCAACCCA
AGCAAGATCTACACACAAGCAGACATAGATACTCTCGATCTCTCCTCGTACGAAAACGAA
CTCCTAGCAGCATGGGATACAGACTCATCCCTTCAACGTGGATTTGACTGGGAGATTGAA
AAGCTACGACGAAATTTTGCAGGCCTGCGCCAAAGAGAAGATGGACAGTGGGTTCGTAAA
CCAAGTCTATTCGATTTCCTCGTCACCAACACACCATCGAACGTAGTTGGTGTGAGCAAT
ACTGGTGAACGATATGAAAGTCCTCCAAAACCGGTGAATATGCTTGATGTAGGATTGCTG
ATCACCAAGAATCTATTGAACACTCTTGGATTTGGGCCGAGTCTCGGTATGGCAGCCGTG
CCGGATGCAGTGATTCAAAAGTATGAAGGAAGTTTCTTCTCCTTCATCAAGGGTGTGCTC
GGTGGTGATCTTCAAACACTCGCAGGAGGACCACTATTCCTACTCCTTGCAAAGTATTAT
CAGGACTATGGACCCATCTTTAATTTGAGCTTTGGCCCGAAGAGTTTCTTGGTCATCTCC
GATCCTGTTATGGCGAGGCATATCTTGAGGGATAGTAGTCCAGAGCAGTATTGCAAGGGA
ATGTTGGCGGAGATTTTGGAGCCAATCATGGGCGATGGTCTTATCCCTGCTGATCCAAAA
ATTTGGAAG

Exon 2:
>Thaps3 chr_9:684641-685198
GTACGACGACGAGCAGTCGTCCCTGGCTTCCACAAAAAGTGGCTCAACAATATGGTGACT
CTCTTTGGTGACTGTGGTGAACGTCTCGTTAACGATCTCGATGCACGGGCTACTGCTAAG
ACTCCAGTGGATATGGAAGAACGATTCTGCTCCGTAACGCTGGATATCATTGGAAAAGCA
GTCTTCAACTATGACTTTGGTTCAGTTACGAAGGAGTCTCCAATTGTCAAAGCAGTGTAC
CGCGTGCTTCGTGAGGCAGAACATCGATCATCATCTTTCATTCCGTACTGGGATTTGCCC
TATGCTGATAAGTGGATGGGAGGTCAAGTTGAATTCCGAAAAGATATGGGCATGTTGGAT
GATATCTTAACAAAACTAATCAATCGCGCTATTGAGACTAGGGACGAGGCAAGCGTAGAA
GAGTTGGAGGATAGAGATGTTGGAGATGATCCAAGTTTGTTGAGGTTCTTGGCTGATATG
AGAGGAGAGGATTTGACGAGTAAGGTGTTGAGGGATGATTTGATGACAATGCTTATTGCA
GGACACGAGACAACGGCT

Exon 3:
>Thaps3 chr_9:685302-685832
GCAATGCTTACTTGGACAGTGTTTGGACTTGTGAGCAATGATTCTGGTTTGATGAAGGAG
ATTCAGGCCGAAGTACGAACAGTCATGGGTGACAAATTGCGTCCAGATTACGATGACATT
GCCAAAATGAAGAAGATGAGATATGCTTTGATAGAAGCACTTCGTTTGTATCCAGAACCA
CCTGTTCTCATTCGTCGGGCAAGGTCTGAGGACAACCTTCCAGCGGGTGGGTCTGGTTTG
TCGGGTGGTGTCAAAGTATTGCGAGGAACAGACATCTTCATTTCTACATGGAATCTTCAT
CGTGCTCCAGAGTATTGGGAGAATCCGGAGAAGTATGATCCCACGCGATGGGAACGACGA
TTCAAAAACCCCGGAGTGAAGGGCTGGAATGGATACGACCCAGAGAAACAATCAGAGTCG
TCGCTGTATCCGAATGAGATCACTGCGGACTATGCATTCCTTCCGTTTGGTGCAGGGAAG
CGAAAGTGCATTGGGGATCAATTTGCAATGCTCGAAGCATCAGTCACTCTG

Exon 4:
>Thaps3 chr_9:685927-686236
GCCATGATCATAAACAAGTTTGACTTTACATTAGTTGGCAGTCCAAAAGATGTCGGCATG
AAAACTGGGGCAACCATTCACACCATGAATGGACTCAACTTGGTGGTGAGTCGTCGGTCG
GAAGATAATCCGATTCCGGAGACCAATGATTACTGGATACAGCAGCATTTGTCGAGAGGT
CTCAATGTCAATGGACGACCATATTCAACCAATGAAGATGCTGCCTGGACGGCATCTTCT

CGAGATAAGAATGAGGGAGTTGTCTCTCGGTTAGTTAATTAAAGTTATCTAGGATATAAG
AGGATTCTGT

Translated:

MCTKLSSRRTLLALYFAFTGCTAFQLPSATPSRASITKAYSTHLDKEIKSKTPLVNP
SKIYTQADIDTLDLSSYENELLAAWDTDSSLQRGFDWEIEKLRRNFAGLRQREDGQWVRK
PSLFDFLVTNTPSNVVGVSNTGERYESPPKPVNMLDVGLLITKNLLNTLGFGPSLGMAAV
PDAVIQKYEGSFFSFIKGVLGGDLQTLAGGPLFLLLAKYYQDYGPIFNLSFGPKSFLVIS
DPVMARHILRDSSPEQYCKGMLAEILEPIMGDGLIPADPKIWKVRRRAVVPGFHKKWLNN
MVTLFGDCGERLVNDLDARATAKTPVDMEERFCSVTLDIIGKAVFNYDFGSVTKESPIVK
AVYRVLREAEHRSSSFIPYWDLPYADKWMGGQVEFRKDMGMLDDILTKLINRAIETRDEA
SVEELEDRDVGDDPSLLRFLADMRGEDLTSKVLRDDLMTMLIAGHETTAAMLTWTVFGLV
SNDSGLMKEIQAEVRTVMGDKLRPDYDDIAKMKKMRYALIEALRLYPEPPVLIRRARSED
NLPAGGSGLSGGVKVLRGTDIFISTWNLHRAPEYWENPEKYDPTRWERRFKNPGVKGWNG
YDPEKQSESSLYPNEITADYAFLPFGAGKRKCIGDQFAMLEASVTLAMIINKFDFTLVGS
PKDVGMKTGATIHTMNGLNLVVSRRSEDNPIPETNDYWIQQHLSRGLNVNGRPYSTNEDA
AWTASSRDKNEGVVSRLVN-

Targeting:

SignalP 3.0 - NN: Yes, Cleavage Site CTA-FQ

SignalP 3.0 - HMM: Yes, Cleavage Site CTA-FQ, Signal Peptide Probability = 0.990

SignalP 4.1: Yes, Cleavage Site CTA-FQ, D = 0.724 (D-cutoff = 0.450)

ChloroP: No, Score = 0.497

10) Thaps3_270370 (ZEP1)

No introns.

>Thaps3 chr_6:1420465-1418423
CTCCGTTCTCATTTTTCCTTTTCTTTCCAAACTTCTTCCGCTGCTGGAGCACCAACGGGT
AAACAGTGGCAGCAGAGGACAAATCAACTAGTACAGTACAACGAATCATGACGGTTAGAA
GAATCGCCTCGCTGGCCATTGGTATCTCGCTGTCCACCTTGACATGTGCATTTGTCACGA
TCTCCTCGTCTCGTACTACCATCAAACCTCTCAATGTCGTTGGCGAACAGGCCTCTTCCA
TCGGACCAGCAACTCTCCTCCGTAATCTCAAACAAAACCTACCACAAATAGATTGGCTAG
CGGAAGGTAAAGGCTCGCCATCCAACAAGATTGACATTCCTGATCATGTAGCTACAGTGC
TTGCTCAACCAAATGCTCCAAAGCGCGAAGCTGAGAGTGAAGAACGTACGCATAAGATCC
GCAGTAGAGCAAAGCAGGCTAGCGAAGATGCAATGGCATTGCGTGGTATGTTGATTGGAG
ATGATGACGCCAATGCATGGTGGCGTGAACAACGTTCTATTCCAGAGGGAGGACGAGTAG
TTACAACGGATGATCCATTGACTGTTCTTGTTGCTGGCGGTGGATTGGCTGGTTTGGTTG
TCGCTGCAGCATGCCATTCCAAAGGCATGAAAGTTGCATTGTTTGAGCAGGCTTCGTCTT
ATGCTCCTTACGGAGGCCCAATTCAGATTCAATCCAATGCATTGAGGGCTTTGCAGCAAA

TCAATCCTGAGATCTTTCAGGAGTTGGTTACTGCTGGAACATGCACTGCGGATCGTGTGT
CTGGATTGAAGATTGGATATAAGAAGGGAAACAAACTTGCTGGACTGTACGATGCAGGAG
ATTGGTTGGTGAGGTTTGACACTATCGGACCAGCGTTGGAAGCTGGATTGCCAGCAACTG
TGGTTGTGGATAGGCCAGTCATTCAGCAGATTCTGGTGAAATATGGCTTTCCTGAGGGCA
CCGTGCGTATCAAATCACGTATCCAATCGTATGAGGATCTGGGGAAGGGACGTGGAGTGA
GTGTCACCTTGGAAGACGGGACGAAAGCGTACGCAGATGTATTGGTAGGAGCTGATGGCA
TCTGGTCTCAAGTTAGAAAGAATCTCCACGGATTGGACGATGGAGCTGGAGGGTTCGCTG
CATCGGGCGCAGCAGGAGGTGCATTGGACGATGCCGAAGCACGCAAATTGGCACGTGATA
CAGTCGCAATTGCAGCCAAGGCCGATCGTCGCTTCTCTGGTTTCACATGTTACGCTGCAC
TGGCTCCTCATCGGGCATCCAACATCGAAAATGTGTCGTATCAAATCTTGTTGGGAGAGA
AGAAGTACTTTGTATCTACCGATGGAGGAGGAGACAGGCAACAGTGGTTTGCACTCATTC
GCGAACCTGCCGGGGGAGTGGATCCTGAGCCCACTCCCGAGGATCCTCACCCTAAGCTCA
CTCGTCTTAGGAAGGAATTTGCGTGCAATGGAAGTGGTGATGCTGATGGCAATGTGTGGG
ATCCATTTGCATTGGAGTTGATCAATGCAGCCTCGGAAGAAGACATCAAGCGTCGTGATT
TATACGACGGAGCTCCTCTCTTGACTACCCTTGACCCACAACGTTTGTTGAGTCCATGGG
CAAAGGGACCTGTGGCACTTTGCGGAGATGCAGCACATCCAATGATGCCTAACCTCGGAC
AAGGAGGATGCCAAGCCACAGAAGACGGATACCGTCTCGTCGAAGAGTTGGCAAAGGTGC
AGCATTCAAGAGATGTTCCAGGAGCACTCGGGAGATACTCTCGCGTTCGTGTGATTAGGA
CAGCCATTATCCAAGGTTTTGCTCAGCTTGGAAGTGATCTGTTGGTTGACTTTGATCTGA
TGATGACTATTCCACTCTTGGGTCCCTTCTTCCTGACAATGACACAGCTTTCCATGCCAT
TCATTTTGAGATACCTCTACACACCTTCTTTTTAAGAGAGATGCTTGTACGATTGTTGAG
AAGCAGATGTTAATGGGTAATATTCTGATAAGGTTTAAACTTTATGATGGAGTTGTCTGT
CCA

Translated:

MTVRRIASLAIGISLSTLTCAFVTI
SSSRTTIKPLNVVGEQASSIGPATLLRNLKQNLPQIDWLAEGKGSPSNKIDIPDHVATVL
AQPNAPKREAESEERTHKIRSRAKQASEDAMALRGMLIGDDDANAWWREQRSIPEGGRVV
TTDDPLTVLVAGGGLAGLVVAAACHSKGMKVALFEQASSYAPYGGPIQIQSNALRALQQI
NPEIFQELVTAGTCTADRVSGLKIGYKKGNKLAGLYDAGDWLVRFDTIGPALEAGLPATV
VVDRPVIQQILVKYGFPEGTVRIKSRIQSYEDLGKGRGVSVTLEDGTKAYADVLVGADGI
WSQVRKNLHGLDDGAGGFAASGAAGGALDDAEARKLARDTVAIAAKADRRFSGFTCYAAL
APHRASNIENVSYQILLGEKKYFVSTDGGGDRQQWFALIREPAGGVDPEPTPEDPHPKLT
RLRKEFACNGSGDADGNVWDPFALELINAASEEDIKRRDLYDGAPLLTTLDPQRLLSPWA
KGPVALCGDAAHPMMPNLGQGGCQATEDGYRLVEELAKVQHSRDVPGALGRYSRVRVIRT
AIIQGFAQLGSDLLVDFDLMMTIPLLGPFFLTMTQLSMPFILRYLYTPSF-

Targeting:

SignalP 3.0 - NN: Yes, Cleavage Site TCA-FV

SignalP 3.0 - HMM: Yes, Cleavage Site TCA-FV, Signal Peptide Probability = 0.960

SignalP 4.1: No, Cleavage Site TCA-FV predicted, D = 0.436 (D-cutoff = 0.500)

ChloroP: Yes, Score = 0.526

11) Thaps3_261390 (ZEP2)

Exon 1:
>Thaps3 chr_2:1117778-1118447
TCCCCCTCTTCGGGTTGTGTCAGACGCAGACGTCCCTCTCGCATTTTGTAGTTCAGTCCA
AGCAACACCAATCAATTAAACAGAATGAAGCTCTCCATCGTGTGCTTCATTATATTAACG
TCGGCGACGTCAGCCTTCATCGCTCCATCAACCACACGCACATCAGTTGCCGTCACAACG
TCATCATTTGCAAATGTGCGAGGAAGTGCCCTCCAAATGGCCGATGACGAAGCCGACGCC
GACTTCAATTCTTCGGATTATGAACTCCTCGGAAGACCAGCTCGCCCTGGGCGTCCTCTC
AAAGTAGCCATTGCTGGTGGTGGTGTCGGTGGTCTCACAGCTGCGCTATGTATGTTGAAG
AAGGGATTTGACGTGACGGTGTACGAAAAGACTGCTGCCTTTGCTCGTTTCGGTGGACCC
ATTCAGTTTGCCTCAAATGCTCTTTCTGTCATCAAGGAGATTGACGAGGAGTTGTTTGAA
CGTGTAATGGATAAGTTCACCTTCACTGGTACAAGAGCTTGTGGTATCAAAGACGGTTTG
AGAGCGGATGGATCGTTCCGTATGACGAATGATTCCTTGGACTACTTGTGGAATCCCGAG
GCTCCTGCTGATTGGTTTGTCAAGTTCCCTTTGAGGCAGTGTGCTGATTTGTTTGGACTT
CCCTACACTG

Exon 2:
>Thaps3 chr_2:1118548-1119207
GCGTCATTGACAGACCCGATTTGCAGGAAATTCTTCTTGATGAGTGCAGAAAGATCAAGC
CCGATTTCATTCAAAATGGCAACCCAGTGAACGGATACGTTAGCAAAGGAAAAGGCAACG
GAGTGACTGTGAACCTCGCCGATGGAACAACTGCAGAGGCTGACGTCCTTGTTGGTTCGG
ATGGTATTTGGTCTGCTATTCGTGCTCAGATGTATGGGGAGGAGATTAAAAAGAGTTCAA
ACAATGCACTCAAACGTCAGGGCTGCACGTACAGTGGATATACCGTCTTTGCTGGAGAGA
CTGTGCTCAAGACGGAGGATTACTACGAGACTGGATACAAAGTGTACATTGGTCCTCAAC
GCTACTTTGTGACTTCAGATGTAGGAGACGGAAGAGTGCAGTGGTACGCTTTCTTTGCCT
TGCCGCCGGGTACGAAGAAAGCACCAAGTGGATGGGGAGGTACCGAGCGAACAGCGCAGG
ACGACCCAGAGGAGAATCTCGTAGATTACATCAAATCGTTGCATCAGGGATGGTCGGATG
AAGTCATGACTGTTCTTGATTCTACCCCTCCTGATAGTGTTGAGCAACGTGACTTGTACG
ATAGGCCACCTGAGCTATTGAGAAGTTGGGCTGATGGAAACGTCGTCCTCATTGGTGATG

Exon 3:
>Thaps3 chr_2:1119282-1119556
CTGTCCACCCGATGATGCCAAACCTTGGACAAGGAGGATGCCAAGCAATTGAAGATGCAT
TTGTTCTTTCTGAAACGCTGGAGGCATGCGAATCTACTCAAAAGTTGGAGGATGCTTTGC
AGGACTTTTACAAAAAGCGTATCGTTCGTGTTAGTATTGTGCAGTTCCTCAGTCGGTTAG
CGAGTGACTTGATCATCAATGCGTTTGATACACCCTGGAGTCCTCATGACGACCTCGGAA
AGTCGTGGAAGAGTTATTTGACTTTCTTCTGGAAG

Exon 4:
>Thaps3 chr_2:1119650-1119989
CCCATTCTTCAGTATGCCATCTTCCCTGCACAGTTTGCTTATCTTTACTCATACCACCCA
ACGGGAAACATGGGAGGTTTGCCTTCTGCTCTTGAAGCAAAGTGGAAGAAACAGCACGAA
GAGGACGCTGAGATGGCATTCAATAGGGTAGAGGAAGAGGGGCAATCAACTAGAGGACCG
AGTTTCTTCAAAATAGCAGAGTCAGAGACGGTGTTGGCTGCAAAGAAGATGTAATGACTA
AGAAGGAAGGAAGAGGGATATGATGCGTGTGTACTCAGGCTACAAGGGGTTGTTAGTGGG
TGAAGAGTGCAACTTATTTAGTACAAAAGTACAGAATGCA

Translated:

MKLSIVCFIILTSATSAFIAPSTTRTSVAVTT
SSFANVRGSALQMADDEADADFNSSDYELLGRPARPGRPLKVAIAGGGVGGLTAALCMLK
KGFDVTVYEKTAAFARFGGPIQFASNALSVIKEIDEELFERVMDKFTFTGTRACGIKDGL
RADGSFRMTNDSLDYLWNPEAPADWFVKFPLRQCADLFGLPYTGVIDRPDLQEILLDECR
KIKPDFIQNGNPVNGYVSKGKGNGVTVNLADGTTAEADVLVGSDGIWSAIRAQMYGEEIK
KSSNNALKRQGCTYSGYTVFAGETVLKTEDYYETGYKVYIGPQRYFVTSDVGDGRVQWYA
FFALPPGTKKAPSGWGGTERTAQDDPEENLVDYIKSLHQGWSDEVMTVLDSTPPDSVEQR
DLYDRPPELLRSWADGNVVLIGDAVHPMMPNLGQGGCQAIEDAFVLSETLEACESTQKLE
DALQDFYKKRIVRVSIVQFLSRLASDLIINAFDTPWSPHDDLGKSWKSYLTFFWKPILQY
AIFPAQFAYLYSYHPTGNMGGLPSALEAKWKKQHEEDAEMAFNRVEEEGQSTRGPSFFKI
AESETVLAAKKM-

Targeting:

SignalP 3.0 - NN: Yes, Cleavage Site TSA-FI

SignalP 3.0 - HMM: Yes, Cleavage Site TSA-FI, Signal Peptide Probability = 0.998

SignalP 4.1: Yes, Cleavage Site TSA-FI, D = 0.793 (D-cutoff = 0.450)

ChloroP: Yes, Score = 0.551

12) Thaps3_7677 (VDE)

No Introns.

>Thaps3 chr_8:842679-841117
AGGACAACAACCAACACCCTTGCCATTGAACACTCTGTTGCTGCTCTACCCAATAGACGC
CATGAAGCTGTTCCTTTCTCTCGTGCTGGCGGCTGCACCAGTGTCTTCCTTCGCACCATC
AAACCCAGTTGTATCTCGTACTCATTCATCGGTACACTCCCAACAACATAATCACGTGCT
CGAAGCTCACAACGACAACATGGATGACATTACCTTCTCCCTATCCGCCAGGAATATCAA
CAACGAGATCGTAGAGCGAATTGGCAAAGTAACCACCTCAGCTCTCCTCGCATTGACGCT
CAGTTTCTCTGCCATCACATCGCCCATCTCCGGTCCCAACGGCGACGTACTATCGTCCAT
CCCATCTGCCAACGCTGCCGATGGTGCAAAGATCGGTCTATGCCTTGTCAAGAAGTGCAG
AGTTCCTTTGGCCAAGTGTATCACCAACCCAAACTGCCTTGCTAATGTGATTTGCATCAA

TTCTTGCAACGGAAAGGAAGATGAGACTGGATGTCAGATTAATTGTGGAAACGTCTTTGA
GAATGACGTTGTTGGGGAGTTCAACAAATGTGCCGTCACCGATATGACGTGCGTTCCTCA
AAAGAAAGACGATGGAAGTTATCCCGTCCCATCCAAAGATGTATTGGTCCAATCGTTTGA
TACCAAACTATGGAACGGAAGATGGTTCATCACCGCAGGGCAGAACAAGCTCTTTGATAC
GTTCCCATGTCAAGTCCACTTCTTCACCGAGACTGCTCCAGGCAAGTTCGTCGGGAAATT
GAATTGGCGTATCGAAGAGCCTGATGGAGAATTCTTCACTCGTGATGCTGTGCAAGAGTT
TGTTCAGGATCCCAACAATCCTGCTCACTTGATCAACCACGACAATGAATATTTGCATTA
CCAAGATGATTGGTACATTGTGGATTATGCCGCGGATGATAACAAGGAGGGTGTTCCTCC
CTTTGCATTTGTGTATTACCGTGGTGAGAATGATGCATGGATTGGATACGGTGGAGCTGT
GGTGTACACTCGTGATTCAAAGTTGCCAGAGTCTCTCCTACCACGTCTTCGTGAGGCTGC
TAAGAAGGTAAACTTTGACTTCGACAAAGACTTTGATCTCACGGACAACTCGTGTAAGGC
ACTTGAGAAGGGAGAGGAGGTCGTGTTGAGGGAGAAGTTTGCTGGTAAGATGGCCATTCA
GACGGAGAAGCAGTTGCAACAGCAGGCTGTGTTGGCACGAACTGCGGCTAGTAATACTGT
AAAGGGTGAAGTGACTGCTGTTGAGAAATCGCTTCAGAAGATTGAAGAGAAGGCTTTGGC
GTTTGAGAAGGAATTGATGAAGGATGTTGTTTCAGTGGAAAAGGAGATCGTAAAGGAAGT
TGAAGAGGTAGAGAAGGAGATTGTTCAAGAGGAACAAAAGATCTTTGGTGGTATTAGATA
GACTGGTTGTAGTTAGTGGAGTAAATGCTACTCATTGACTTTATCTTGGAATGATATTTG
AGAAGACGTCTAAATGCACTGCGAGTAAGAAGATCTTTATGTGAATGGGAAAAGTATGAA
TGG

Translated:

MKLFLSLVLAAAPVSS
FAPSNPVVSRTHSSVHSQQHNHVLEAHNDNMDDITFSLSARNINNEIVERIGKVTTSALL
ALTLSFSAITSPISGPNGDVLSSIPSANAADGAKIGLCLVKKCRVPLAKCITNPNCLANV
ICINSCNGKEDETGCQINCGNVFENDVVGEFNKCAVTDMTCVPQKKDDGSYPVPSKDVLV
QSFDTKLWNGRWFITAGQNKLFDTFPCQVHFFTETAPGKFVGKLNWRIEEPDGEFFTRDA
VQEFVQDPNNPAHLINHDNEYLHYQDDWYIVDYAADDNKEGVPPFAFVYYRGENDAWIGY
GGAVVYTRDSKLPESLLPRLREAAKKVNFDFDKDFDLTDNSCKALEKGEEVVLREKFAGK
MAIQTEKQLQQQAVLARTAASNTVKGEVTAVEKSLQKIEEKALAFEKELMKDVVSVEKEI
VKEVEEVEKEIVQEEQKIFGGIR-

Targeting:

SignalP 3.0 - NN: Yes, Cleavage Site VSS-FA

SignalP 3.0 - HMM: Yes, Cleavage Site VSS-FA, Signal Peptide Probability = 1.000

SignalP 4.1: Yes, Cleavage Site VSS-FA, D = 0.682 (D-cutoff = 0.450)

ChloroP: Yes, Score = 0.529

13) Thaps3_22076 (VDL1)

Exon 1:
>Thaps3 chr_4:141771-142931
CCCATAACAGCGCTTCCAACACCGCAACATGAGACCGTCAACCTCTGCCTTGACAGTCGT
CCTAGGCACCATTGCGCTTGTCAGCTGCAGCCAGCTCAACAACAATGTATCCGCCTTCTC
TACAAGATCATCGTCGTTGACGCAAAGACACAAGACATGCACCATCACAACATCATCATC
ATCTCTCTACGTGAACCCCAACAACGATGATGACAACTCCAACAGATCCCAACACAAACC
CAATCCCTTCCTCTCAGCGGCACTAACCGCCGCAGTCACCACATCCCTCTTCCTCTCATC
TCTCCCCAGTGCCACCTTCGCCTCCACACCGGCCTCCACAACGCAAAAGTACGACGGCTT
CGCCGAGTACGCCAAGGAAAACAAAATGGAACAATCGGACGTAGGATGCTTCATTAACAA
GTGTGGCGATCAGACGAAACAACTGTTTAGTAATCCCCGTGGTATCAAGGGGGTGTCGTG
TTTGGGACGGTGCAAGGGGGAACAATCGTGTGCTACGCGGTGTTTTGCTGAGTTTGGGAG
CGAGGATTTGGACAACTGGTTATCGTGCACTATTGAAGATTATGAATGTGTGAAAGTTCC
AAAGAATATTGACAACTCTGCCGAGAATGTGGGGTATGATACTACCGTGAAGAAGTTTGA
TCCGTCAACGTTGGTGGGAAAGTGGTACAAGACGGATGGACTGAATCCCAATTACGATCT
GTTCGATTGTCAATCTAATACGTTTGACTTTTCAGATGATACGAAAAAGGAGTTGGATAT
GGGTATCTTCTTTAGAGTGCCGCGTCCAGAAGAATACGGAGGTGGATTCTGGGAGAACAG
TCTTACAGAGCACATGATTGTTGATGCCGTATCACCCGAGTTAGACAACCCTACTGGAAG
AACGATGCATACCGCTGGTAAGATGTATGGGCTCAAATTCACTGAGAACTGGTACATACT
CGGAGAATCCAATGGTGATAATGATATCCCTCCGTTTAAGTTGGTGGCGTATAAGGGGCA
TACGTTGCAAGGGAATTATGAGGAGGCGTTTGTGTATGCGAAGGAGAGTGTTTTGCCGAA
GGAGGCTGTGGGGGCGGTGAGGGAGGCTGCGGCGAAGGCTGGACTGGACTTCGACAAGTT
CACGAGGATTGATAATACTTG

Exon 2:
>Thaps3 chr_4:143069-143272
TCCAACCACAACAAAATCACTGAATGATGCATCGGCTGGAACTGGAACGTCTACCACAGA
CTGGGTTGATCTTGTAGTTGGTGAAGGAGGAGTTATTGATTGGGTTGTTCCTGGATGGAG
AGGAGAGTACAAAAACTAAAACAGGTTAGAGCGTCGAAGAAGAAGATGCTTTTCTAACCT
AAAATCTGGGTATTGTCGTCCCGA

Translated:

MRPSTSALTVVLGTIALVSCSQLNNNVSAFSTRSSSLTQRHKTCTITTSSS
SLYVNPNNDDDNSNRSQHKPNPFLSAALTAAVTTSLFLSSLPSATFASTPASTTQKYDGF
AEYAKENKMEQSDVGCFINKCGDQTKQLFSNPRGIKGVSCLGRCKGEQSCATRCFAEFGS
EDLDNWLSCTIEDYECVKVPKNIDNSAENVGYDTTVKKFDPSTLVGKWYKTDGLNPNYDL
FDCQSNTFDFSDDTKKELDMGIFFRVPRPEEYGGGFWENSLTEHMIVDAVSPELDNPTGR
TMHTAGKMYGLKFTENWYILGESNGDNDIPPFKLVAYKGHTLQGNYEEAFVYAKESVLPK
EAVGAVREAAAKAGLDFDKFTRIDNTCPTTTKSLNDASAGTGTSTTDWVDLVVGEGGVID
WVVPGWRGEYKN-

Targeting:

SignalP 3.0 - NN: Yes, Cleavage Site VSC-SQ

SignalP 3.0 - HMM: Yes, Cleavage Site VSC-SQ, Signal Peptide Probability = 0.970

SignalP 4.1: Yes, Cleavage Site VSC-SQ, D = 0.694 (D-cutoff = 0.450)

ChloroP: Yes, Score = 0.568

14)Thaps3_11707 (VDL2)

Exon 1:
>Thaps3 chr_22:194840-195897
ATTTCCTTCACTTTTACTTTGATTCAAATCTTAATAGACATAGCCTTTTTACGTGGCAAA
GGTCGCGTCATTTGCCAGGAGGTTGAAGCAAAACAGTGGACGACGTCGCTCGTGAGGACG
GACAGCGAGAGAGGCCACCCCAGCTGTTGCATCCCAACAACCATCCATCGTAGACTTCAA
TACACATCAACTGATAACAGAGTGGTGCTGATACTGTCAGTTAGAGTTCATTGGATTAGA
AAGACTGCAAGTGAGATACACCACAGTCCAGATCGCCTGAGTTGTAATAGTCAAAGTTTT
TGCTCCCTCAATACAAAACAAAATGTCGGCATCATCATCATCAACAACAACAACAACAAA
CGCTGGAAAGAGGGCACGATCGTGGCCGTCTTCATCCGCATCAACATCATCAATGCCTAC
TCGGTCGATACTGATACTGGCAACCTTTCTCTCATTGACGTCGTCGTCGTCGTCAACATC
GGTGGAAGCTGCCTTTGTCGGCAGCCCTGCCGTTGGACTTAGATCTCACACTGCTGCGTC
AACATCAAAGCAGCAGTCGTCTCTCTACGCTCAAAAGAAGAACAACATCGACTCATCCGA
CAACCCGCTATCATATCTATTCGATCTATCCTCCGATCCAGAAACCAAACGTCGCCTCCA
AAAACAAACAGCCACCCTCTTCTCCACCCTCGGCTTCTCCGCCCTCTTCCTCACCAATCC
ACTCATCCCACACCTCCCCTTCTCCCCCTCCCTCAGCAGTGCCAACGCCGAAGACGAACT
CTATGCTAAATATGGTGGCAAAGGCCTCGATACATCCTTAGTCGACAAAGACTGTCTCGT
CAATCAATGTCAAGTGCAAGCCAAAGCGTGTCTTCAGGATGATCCCGATTGTAGAAAGGG
ATTAACGTGTACTGCCAAGTGTTTGGGGGATAATGCGTGCATTACGGGGTGTTTTGCTAG
GTATGGGAATGAGAATTTGGATGAGTTGTTGAAGTGTACTATTGAGGATCATGAGTGTAT
TAAGGTGGCTATTTTGGAGGGTGGGGGCGATGTGCTTG

Exon 2:
>Thaps3 chr_22:195996-196062
GGCGAGAGCCAAAGTCGCCTGCTCCTACTGTTCAAGGGTTTGATCTAGCCAGTATGGAAG
GGACTTG

Exon 3:
>Thaps3 chr_22:196169-196611
GTATAAAGTAGCAGGCTACAACCCCAACTACGATTGCTACGCCTGCCAACGAAACACCTT
CTCCTCACCCGAAGGCGGTCTCTCCGACTCACTCCAACTACCAACAGGAGGAATACTCGG
CTCTCTATCCAACGCTGTAGGTTCCATCGGTGCAGATCGCTTACAAGTCGATGTGGAATT
CAGTATGCCGAGGTATTTGCCAGATGGTAGTCCTCAGCCACCGAGTGGAGTGCGCGAATC
ATTCATTAGTAGTGCTGATTCAATGGAAGGGAGTGGGTTGCAGAGTGTGGGGTACAATCA
GTACTCGACTCATGAGACAATGGTGTTTGATACGGTCAAGAGTAATGGTGTGGGGGAAGC

TGTGAAATTGGCTTTGGGTAAGAGGGGAGAGGAGAAGTTGTATTCGAGGACGGCTCATTC
GGAAGGAGAGATGTTTGGACTGA

Exon 4:
>Thaps3 chr_22:196740-197411
AATTCTGGGAGAACTGGTACATCATTGGCCAAAACAACCCCGGCCAAGACGAGTTCAAAT
TCGTCTACTACAACGGCAAGACACGTCAAAATACCTACGACGGTGCCTTCATCTACTCCC
GATCACGCACCCTCTCACCCGCGTCCATGGAGAAAGTCTACAAGATTGCCAAGGATGCGG
GTATGAATCCTGATCAGTTTTGTAAGATTCAGAACTCTTGCTTTGACGGTGAGGATGATA
AGCAAGAGATGATGATGATGAATCCTCAGAGAGAAGGACTGGGTAGTCCTTCCAATCCAT
TTAGGGGTATTTTGGCATCTACGAAAGTATCTCAATTCTTGGGAGTTGAATCTGTGGCGG
CCGAGACTACGTACAACGAACCAAAGAGTACGATATCGTCCAACTTTCTTCAAGGGAGTC
AGGCTACCAACCGCAAAGCGGATGCCGTTCAGGAGAGGCCGTGGTGGAAGGAAATGGGAG
ATTATTTGGAAGATCCTAGACGGCATTTCCGTTTGATGGATAGTCTGAGAACAGATATGG
ATTGGCCGGATTATATCAAAGAGAAGAATTGGTGAAGGAGTTCGGCATGCTGAGTAGTTA
CGAAACTACTTTTGTTTGGAGTATGTGTACTATTTCTATGATCCAAAGATGTAAGGGGTA
ACTTAAAGAAAC

Translated:

MSASSSSTTTTTN
AGKRARSWPSSSASTSSMPTRSILILATFLSLTSSSSSTSVEAAFVGSPAVGLRSHTAAS
TSKQQSSLYAQKKNNIDSSDNPLSYLFDLSSDPETKRRLQKQTATLFSTLGFSALFLTNP
LIPHLPFSPSLSSANAEDELYAKYGGKGLDTSLVDKDCLVNQCQVQAKACLQDDPDCRKG
LTCTAKCLGDNACITGCFARYGNENLDELLKCTIEDHECIKVAILEGGGDVLGREPKSPA
PTVQGFDLASMEGTWYKVAGYNPNYDCYACQRNTFSSPEGGLSDSLQLPTGGILGSLSNA
VGSIGADRLQVDVEFSMPRYLPDGSPQPPSGVRESFISSADSMEGSGLQSVGYNQYSTHE
TMVFDTVKSNGVGEAVKLALGKRGEEKLYSRTAHSEGEMFGLKFWENWYIIGQNNPGQDE
FKFVYYNGKTRQNTYDGAFIYSRSRTLSPASMEKVYKIAKDAGMNPDQFCKIQNSCFDGE
DDKQEMMMMNPQREGLGSPSNPFRGILASTKVSQFLGVESVAAETTYNEPKSTISSNFLQ
GSQATNRKADAVQERPWWKEMGDYLEDPRRHFRLMDSLRTDMDWPDYIKEKNW-

Based on the targeting analysis of the above peptide (a delayed signal sequence and clear chloroplast targeting), it appears that the protein starts at the second in-frame methionine. The following analysis pertains to the latter.

Targeting:

SignalP 3.0 - NN: Yes, Cleavage Site SSS-TS

SignalP 3.0 - HMM: Yes, Cleavage Site VEA-AF, Signal Peptide Probability = 1.000

SignalP 4.1: Yes, Cleavage Site SSS-TS, D = 0.799 (D-cutoff = 0.450)

ChloroP: Yes, Score = 0.554

15) Thaps3_270211 (VDR)

Exon 1:
>Thaps3 chr_2:1186955-1187856
GGGGGTGATGAGGAGCCGAGGGGCGCCGAGCAACGACGCGAGCAACTACACAACGAGCAA
CCCCACAACAAGAACAATGGCAATGGTGCTGCTGATACGGACAGCAGTGATAGCTTCATA
TAGCCTAACATTGACATCGGCATTTTCCTCATCAATAAGGCCGACTTGCCGGACATTTCG
TCAGAGTACACCACATCATGCAACCGCTTCCAACGTCGACATCATCGGTACAGTGGCACT
TCTCGTGCCTTCTTCGTCCACTGAGCTGTCAAAGTATGGATCCAAATCTCCAGCACCTCG
ACCATCCTATCAAGAAGCAGCTGAACACTTGGCTCGCAAAATAAGTCACTTCTCTGACGG
TCGAATAGAAGCAACAGTAGTAACACCATCAACTAATCAAGATGACACGGACGACGTCTG
CCTTACATCTAATGCACTGATTGCTTTGGGAATCACTGACCCAGCGGAAGTCCAATACCT
TTCCACGACGTTTCGTAAACGTCGCACATCTCATCAAGAAACGTCCTCTTATAACACGTG
TCAATTTGCATTAGACTGCGGCAGCAACAACTATGCACCTCTCGTTGGACCATGGGATGA
AGCCAATCCATCCATTCTCGCCGAAATTGCTCCGTGGACAGGTGTGGCATCTGGCAAACG
TCTGACAGAACAAATGAATGGCTTGTTTGAAAAGCAAACATCTGATGAGTTTGCCTTGGC
AGTTATGCTCTTTTTCAATCGGTTTTCTGGCGCTGCTATTCCTTGGGTGCAACACTCAAT
TGATGTAACTTGGGAGAAGGGATTGGTCCAGAATGCAAAGGAGATCTTCTCTATGATTAC
AAAGTGTGGACCTTGCATTACCAAGTGTTTGAACGATGAGAATTGTTCTCAGTGTATCAA
CG

Exon 2:
>Thaps3 chr_2:1187934-1188181
CACTTGACAAGATCGACACACGAGACCAAGTCACAAGTTATAGAACAGTCGTGTCGTTTG
AAAGTGAACTGCTTAGGGATTTCAGCTTGTGTATTTTGCAAAAGAATAACATCTTCGAGT
GCTCAGCTGAGATCCCAGAGTTGCCAGTTGTCAAACCGATGAGCACATGGAGAGGGAAGG
ATGTTACGACGGACGTCGCGAGAGGCATTATGATTGGGCACTTAGAAGGAGCGGGGGGAT
CGCTAGAG

Exon 3:
>Thaps3 chr_2:1188256-1188336
GGGAATTTGCAACTTGGTGTCTCTTGGAAGGTGGCTTGCGGAGCAAACGTTGCCTATGAT
CAGTTTCCCTCTCAAAATCAG

Exon 4:
>Thaps3 chr_2:1188422-1189161
TTGTTTTACCCATCTGCAAAGGGGAAGGATCTTTGGTACGACCCTGTATTCCGAGTAGAA
ACAATTGACGGTAGAAATGTCTGGTGCAAACGTCACTACAAAGTCAGGAATGGAGAAACT
CCTGGTACGTTCAAATTCTCGGTATTGGACAATGGCGTAACGAGCAACGAGTTCTGGACG
ATTGTCGGGGCTGCTGACGACTTGTCGTGGGTTGTATTTCATTACGCTGGAGCTGCTGGC
GCTGTAGGCCAGAGGTATTTGGGAGGACTGCTGTGCACACCAACGGGAGAGCTGCCACCA
GAAGAAGACCTTGGACACATCTACAATTTTCTTCGATCGGCGGAAATTGAACCGTGGGAG
TTATTTGTTGTGGACAATGATGATCAGTCGCCTGGTGCGTTAGCAGCAGGCGCTCCACCG
CTGGATTACTTCAGAAAGACTGCATCGGTCATTGGCTGAGCATTCGCCTGTACTATGTTT
AATCCTTAAGCCAAGATACTCTCAAGCAACGGCGAATGCTCCGTCTTGCGAAAGTTATGC

ATTACACCCCAGTGCTCATCGTCGAACCGAAATATCACCGCAACGTCAAGAGGCAAGTCG
CTAATAAAAGATGCAGAAGGCAACTGACAAAGCAGCTTTGATCCTGGGAACGTCGCCTTT
TTTGTAACACTTGCCAAGCTTGTAATTCTGCCGACACGTCCGTTGAGGAATGGTAACAAA
GGAACCTCATTTGACTCGTA

Translated:

MRSRGAPSNDASNYTTSNPTTRTMAMVLLIRTAVIASYSLTLTSAFSSSIRPTCRTFR
QSTPHHATASNVDIIGTVALLVPSSSTELSKYGSKSPAPRPSYQEAAEHLARKISHFSDG
RIEATVVTPSTNQDDTDDVCLTSNALIALGITDPAEVQYLSTTFRKRRTSHQETSSYNTC
QFALDCGSNNYAPLVGPWDEANPSILAEIAPWTGVASGKRLTEQMNGLFEKQTSDEFALA
VMLFFNRFSGAAIPWVQHSIDVTWEKGLVQNAKEIFSMITKCGPCITKCLNDENCSQCIN
ALDKIDTRDQVTSYRTVVSFESELLRDFSLCILQKNNIFECSAEIPELPVVKPMSTWRGK
DVTTDVARGIMIGHLEGAGGSLEGNLQLGVSWKVACGANVAYDQFPSQNQLFYPSAKGKD
LWYDPVFRVETIDGRNVWCKRHYKVRNGETPGTFKFSVLDNGVTSNEFWTIVGAADDLSW
VVFHYAGAAGAVGQRYLGGLLCTPTGELPPEEDLGHIYNFLRSAEIEPWELFVVDNDDQS
PGALAAGAPPLDYFRKTASVIG-

Targeting:

SignalP 3.0 - NN: No

SignalP 3.0 - HMM: No, Signal Peptide Probability = 0.163

SignalP 4.1: No, D = 0.173 (D-cutoff = 0.500)

ChloroP: Yes, Score = 0.579

Based on the clear predicted chloroplast targeting, the analysis was repeated with the peptide starting with the second in-frame methionine, with the following results:

SignalP 3.0 - NN: Yes, Cleavage Site TSA-FS

SignalP 3.0 - HMM: Yes, Cleavage Site TSA-FS, Signal Peptide Probability = 0.975

SignalP 4.1: Yes, D = 0.654 (D-cutoff = 0.450)

ChloroP: Yes, Score = 0.561

Based on the clear ER and chloroplast predicted targeting, it is likely that the peptide begins at the second in-frame methionine.

16)Thaps3_bd_1474

No available RNA-seq data for the unmapped “bottom drawer” sequences.

Gene model predicted by JGI (below) cuts off abruptly on the N-terminal side due to the way the unmapped sequences were assembled. Some N-terminal sequence is likely missing.

JGI-Predicted Exon 1:
>Thaps3 bd_37x91:13831-12903
CTTATGAATCAAAAGATGTTGACGTTGGGTGAAAAAATTCAGACCGCTCCTCCTCTTCTTCCTATGCT
TATTGAGGGACAGTCATTCATTGATGCTCAGGATGAGTTGAGTGTGACGCAGTTCATGAGGAAGTACGGT
ATGCCTGAGAGAATCAACGAGGAGGTGTTTATTGCGATGGCCAAGGCGTTGGACTTTATTGATCCTGATA
AGTTGAGTATGACTGTGGTGCTTACGGCTATGAACAGGTTCTTGAATGAGAGTAATGGACTTCAGATGGC
ATTCTTGGATGGAAATCAGCCTGATAGGTGGTGCACTCCCACCAAGGAGTATGTGGAAGCACGCGGAGGA
AAGGTCAAATTGAACTCTCCCATTAAGGAGATTGTGACCAACGACGATGGAACTATCAATCACCTTCTCC
TTCGATCTGGCGAGAAGATTGTGGCCGATGAATACGTCTCTGCCATGCCCGTGGACATCGTCAAACGTAT
GCTTCCCACAACGTGGCAGACTATGCCCTACTTCCGTCAGCTTGACGAACTTGAGGGCATCCCTGTTATC
AACTTGCACATGTGGTTCGATCGTAAGTTGAAAGCAGTCGACCATCTTTGCTTCAGTCGCTCCCCACTCC
TTTCCGTCTACGCTGACATGTCCGTCACATGCAAGGAGTACGAAGATCCCAACAAGTCCATGTTGGAATT
GGTCTTTGCTCCCTGCTCTCCTATTGCCGGAGGAAATGTCAACTGGATTGGAAAGTCAGATGAGGAAATC
ATTGATGCTACCATGGGTGAGCTTGCTCGCCTTTTCCCTACCGAGATTGCGAATGATGATAAGTGGCCTG
CTACGAAGATGCAGGGACCTAATGGACAGGCAAAGCTTGAGAAGTATGCTGTTGTGAAGGTGCCAAGGAG
TGTGTATGCTGCCATTCCTG

JGI-Predicted Exon 2:
>Thaps3 bd_37x91:12821-12475
GACGTAACAAATACCGCCCCAGTCAGACCTCCCCCATC
CCACACTTCACCATGGCTGGATGCTATACCTCACAAAAGTTCCTCGGATCCATGGAGGGTGCCACCCTCG
CCGGGGAAGCTTGCTGCCGAGGTCATTGCCAACCGTGCCCTCGGAAATGCGGATAAGCCAGTCAAGGAGAT
TCAGCAACACATTATCGACTCGGCTAGTAAGCATGTTGTGAAGGAGCCAGTGGGTGTGAAGGGAGAGGGA
GCGATTGCATTTGGAGGGGGGTATACTGTTGGAAAGAAGGAGGAGGATTTGTTGAGGGAGTCGGATCCTG
CTCAGTATGAGTTGGCAGTAGCCAAGTAA

Translated:

MNQKMLTLGEKIQTAPPLLPMLIEGQSFIDAQDELSVTQFMRKYGMPERINEEVFIAMA
KALDFIDPDKLSMTVVLTAMNRFLNESNGLQMAFLDGNQPDRWCTPTKEYVEARGGKVKL
NSPIKEIVTNDDGTINHLLLRSGEKIVADEYVSAMPVDIVKRMLPTTWQTMPYFRQLDEL
EGIPVINLHMWFDRKLKAVDHLCFSRSPLLSVYADMSVTCKEYEDPNKSMLELVFAPCSP
IAGGNVNWIGKSDEEIIDATMGELARLFPTEIANDDKWPATKMQGPNGQAKLEKYAVVKV
PRSVYAAIPGRNKYRPSQTSPIPHFTMAGCYTSQKFLGSMEGATLAGKLAAEVIANRALG
NADKPVKEIQQHIIDSASKHVVKEPVGVKGEGAIAFGGGYTVGKKEEDLLRESDPAQYEL
AVAK-

Genomic sequence disregarding the predicted intron:

>Thaps3 bd_37x91:13831-12475
CTTATGAATCAAAAGATGTTGACGTTGGGTGAAAAAATTCAGACCGCTCCTCCTCTTCTTCCTATGCT
TATTGAGGGACAGTCATTCATTGATGCTCAGGATGAGTTGAGTGTGACGCAGTTCATGAGGAAGTACGGT
ATGCCTGAGAGAATCAACGAGGAGGTGTTTATTGCGATGGCCAAGGCGTTGGACTTTATTGATCCTGATA
AGTTGAGTATGACTGTGGTGCTTACGGCTATGAACAGGTTCTTGAATGAGAGTAATGGACTTCAGATGGC
ATTCTTGGATGGAAATCAGCCTGATAGGTGGTGCACTCCCACCAAGGAGTATGTGGAAGCACGCGGAGGA
AAGGTCAAATTGAACTCTCCCATTAAGGAGATTGTGACCAACGACGATGGAACTATCAATCACCTTCTCC
TTCGATCTGGCGAGAAGATTGTGGCCGATGAATACGTCTCTGCCATGCCCGTGGACATCGTCAAACGTAT
GCTTCCCACAACGTGGCAGACTATGCCCTACTTCCGTCAGCTTGACGAACTTGAGGGCATCCCTGTTATC
AACTTGCACATGTGGTTCGATCGTAAGTTGAAAGCAGTCGACCATCTTTGCTTCAGTCGCTCCCCACTCC
TTTCCGTCTACGCTGACATGTCCGTCACATGCAAGGAGTACGAAGATCCCAACAAGTCCATGTTGGAATT
GGTCTTTGCTCCCTGCTCTCCTATTGCCGGAGGAAATGTCAACTGGATTGGAAAGTCAGATGAGGAAATC
ATTGATGCTACCATGGGTGAGCTTGCTCGCCTTTTCCCTACCGAGATTGCGAATGATGATAAGTGGCCTG
CTACGAAGATGCAGGGACCTAATGGACAGGCAAAGCTTGAGAAGTATGCTGTTGTGAAGGTGCCAAGGAG
TGTGTATGCTGCCATTCCTGGTGAGTGAAAAAGAGTGTGCGGTGAATCTTCGTCGTCTACCTTGGCCAAC
TACTGACTCTTGTTCTCTTTAAATCATATCAGGACGTAACAAATACCGCCCCAGTCAGACCTCCCCCATC
CCACACTTCACCATGGCTGGATGCTATACCTCACAAAAGTTCCTCGGATCCATGGAGGGTGCCACCCTCG
CCGGGAAGCTTGCTGCCGAGGTCATTGCCAACCGTGCCCTCGGAAATGCGGATAAGCCAGTCAAGGAGAT
TCAGCAACACATTATCGACTCGGCTAGTAAGCATGTTGTGAAGGAGCCAGTGGGTGTGAAGGGAGAGGGA
CGATTGCATTTGGAGGGGGGTATACTGTTGGAAAGAAGGAGGAGGATTTGTTGAGGGAGTCGGATCCTG
CTCAGTATGAGTTGGCAGTAGCCAAGTAA

Peptide sequence disregarding the predicted intron:

MNQKMLTLGEKIQTAPPLLPMLIEGQSFIDAQDELSVTQFMRKYGMPERINEEVFIAMA
KALDFIDPDKLSMTVVLTAMNRFLNESNGLQMAFLDGNQPDRWCTPTKEYVEARGGKVKL
NSPIKEIVTNDDGTINHLLLRSGEKIVADEYVSAMPVDIVKRMLPTTWQTMPYFRQLDEL
EGIPVINLHMWFDRKLKAVDHLCFSRSPLLSVYADMSVTCKEYEDPNKSMLELVFAPCSP
IAGGNVNWIGKSDEEIIDATMGELARLFPTEIANDDKWPATKMQGPNGQAKLEKYAVVKV
PRSVYAAIPGE-

No predicted targeting to the ER or chloroplast for either peptide, however due to the high likelihood of a missing N-terminus, such analysis is inconclusive.

17)Thaps3_21900

No introns.

>Thaps3 chr_3:1688852-1690753
CAACGACCGACGGCACACAACCCCAACACCAAAATGATCCGTTCAATATCAGCGTTGGCA
CTCCTGGCAGCCTGCTGCCCATCCGTTTTTTCCTTCGCTCCCCTATCAGTATTTAGAGCA
AATGCACCATCTTCACTGGCATCGACCACCTATCAAGATGAAGTAGACTGCATCGTTATC
GGCAGTGGCATCGGTGGTCTCTCGTGTGCAGCCCTCTTAGCAGCCACAGGTCGTACCGTA

CGCGTCCTAGAGCAACATTATGAAATAGGAGGATGCGCACATGCATTTTACATGGATATG
AATGGCAAGACGGTACCTTCGTCTGCGCTGAAGGATGACCCTACAAAGAAAGGAGAGTTG
TTTCATTTCGAAGCAGGGCCTAGTTTGTACAGTGGACTATCGGAGGAGAGAACACCGAAC
CCGCTTAAGCACATCTATCAAATGATCGAGGAAGAGCCAGAGTGGTTAACGTACGATCAG
TGGGGTGCCTTCTTGCCTGAAGCCCCAGAGGGATATCAGATGAGTATTGGAGCAGAAAAC
TTTTGCAAGATCCTCGAGACTTATGGAGGCGAGGGCGCAGTTGAAGATTGGGAAAAGCTT
GCCGAGCAATTGAGACCAATGGCGGGGGGTATCAAAGGCATTCCACATGCAGCAATACGC
GGTGATTGGGGTATCTTCCTGACTCTCATCTTGAAGTATCCTCTCTCCTTCATGAATGTG
CTCAAGTATGCTCCTGCATTCACTGCTCCTTTTGATTTGGACAAGTTGGGCGTGACAAAC
AAGTTCTTGCGGAATTATCTTGAAATGCTGGCTTTCCTTTTGCAAGGCTTGCCAGCTGAT
CAAACATTGACCGTAGTCATGGCGTATATGGTAGAGGACTTCTTTCGTGAAAATGCAGTG
ATGGATTTCCCAAAAGGAGGATCTGGTGAGCTTATGGGTGCTTTGGCGAGAGGAGTGACA
AAACGTGAGGGGTGTTCCGTCGAAGTATCTACATCTGTGGACGAAGTTATTGTTGAGAAT
GGACGTGCTGTTGGTGTGAAATTGGCAAAGAGCGGACGTATCATCAAAGCGAAGGAAGCT
GTTATCAGCAATGCTGATCTTTACAACACATACAAGTTTGTTCCAGAGGGAAAGCACGAG
GGATTCGACAAAGAGAGAATTGAGTACCTTGGTCTTACTGCAAAGCCGAAAGATGGCTCA
GTTCCATTCTGCAAATCATTCATGCACTTGCATCTCGCTGTAAAGGCAGAACTCATACCA
GAAGATGCGCCTCCACAATGGACTGTTGTTCAGGATTGGGATAAAGGTATCGATGCAACT
GGAAACGTAGTGGTTGTATCGGTCGGAAGCAAACTTGATCAGTCATTAGCACCGCCTGGA
TATCACGTTATCCATGCTTACACAGCGGGAAATGAATCTTACGAAGACTGGGAACAATTT
GAACATCTGATGGATGATGCTGCCGTCAGAGACAAAGATGCAGCATACCAAACATTCAAA
GACGAGCGAGCTCAGCCGATTTGGGATGCCATCCAAAAGCGTGCCCCTGCCGTCGTCAAG
GGTGCTTGTGTTATAGAAAAGGTAGCCACCCCATTGACTCACGCTCGATTCCTCAATAGA
CATCGTGGAAACTACGGATTAGCCATTGCGCCGGATAATGCAGAAGGCTGGAAGTTCCCA
GATGTAAAGACGCCTCTTGAAGGATACTACAGGTGCGGTGATTCCACAACGTCTGGCATT
GGAGTTCCGGCGACGGCAAGTAGTGGAGCCGTTTGTGCTAATGCGATCATGTCCGTTTGG
GATCAGCTTTCATTGAATCAAAAGATCAAAATGCCGTGAACGAGTAAGTATAATCAATCG
TTCTTTCTCTGTAATATATTTAAACTTCAAGCAGCTCAATAC

Translated:

MIRSISALALLAACCPSVFSFAPLSVFRANAPSSLASTTYQDEVDCIVI
GSGIGGLSCAALLAATGRTVRVLEQHYEIGGCAHAFYMDMNGKTVPSSALKDDPTKKGEL
FHFEAGPSLYSGLSEERTPNPLKHIYQMIEEEPEWLTYDQWGAFLPEAPEGYQMSIGAEN
FCKILETYGGEGAVEDWEKLAEQLRPMAGGIKGIPHAAIRGDWGIFLTLILKYPLSFMNV
LKYAPAFTAPFDLDKLGVTNKFLRNYLEMLAFLLQGLPADQTLTVVMAYMVEDFFRENAV
MDFPKGGSGELMGALARGVTKREGCSVEVSTSVDEVIVENGRAVGVKLAKSGRIIKAKEA
VISNADLYNTYKFVPEGKHEGFDKERIEYLGLTAKPKDGSVPFCKSFMHLHLAVKAELIP
EDAPPQWTVVQDWDKGIDATGNVVVVSVGSKLDQSLAPPGYHVIHAYTAGNESYEDWEQF
EHLMDDAAVRDKDAAYQTFKDERAQPIWDAIQKRAPAVVKGACVIEKVATPLTHARFLNR
HRGNYGLAIAPDNAEGWKFPDVKTPLEGYYRCGDSTTSGIGVPATASSGAVCANAIMSVW
DQLSLNQKIKMP-

Targeting:

SignalP 3.0 - NN: Yes, Cleavage Site VFS-FA

SignalP 3.0 - HMM: Yes, Cleavage Site VFS-FA, Signal Peptide Probability = 0.998

SignalP 4.1: Yes, Cleavage Site VFS-FA, D = 0.663 (D-cutoff = 0.500)

ChloroP: Yes, Score = 0.517

18)Thaps3_25361

Exon 1:
>Thaps3 chr_18:367805-365028
CTCGCGAAATGAACCTCGTGATCCTCTTCGTCAAACCACCAACGCCGAACACAAGCACAC
CGTATTGACATATCCAATACGCTACGCGCTACCAATAGAGCTATACTACTAGACGTATCT
TGAACCGTCATCATCAAGCAGCTACGCTAACACGAGACACATCGCCAACACACAGGACAC
CATCCATCATGCGTATGGGACGCCCCAACAAAAAGCTCCGCTCCACCTCCAAGCAAACCA
CCAATCCCAATCCACCCAAATACTCCTCGCCCACATTAGTAGTAGGACAGGTGTCCTCCA
ACGTCATCCATTCCATCTATGGGCCGGCGTTGACGAAGCTGGCCGTGGAGAGCGTGGAGG
AGTATGCAGATGCTGTTTTGAGGTGGGAGGCGAGTTTGCCGGAGGTACTTGTCAAGCCAT
CACAGCTTGATGATGCAGAGGATATTGATGTTGATGCCGATGGCACCTTCGAAAAGGGAG
AGGAGGTGGAAGTCGACCTCGACGGAAGCATCCTGCCCAGCCACGACAACGACGACGACA
AGACATCTTCACCCACAATGCCCACCTCCAACCGCCTCTTCACCACCCAATCATCCATCG
ACAACCTCACCGCCCTCCTCACGGACACATCCCAACACTTCTCTACCACCAACGCATGGA
AGATACACGCCAACGCCGCCAAATTCGAACGTCTATTAGATGAAAAATACGGACGTTTCC
GTCCCTTTATTGAATCTCATCCAGAGTTAGAGGTGTTTATCAAAAAGGTTCAAAGGAAGT
ATGCCATGGGACAATTCAGCCCCCTTAGGAAAGGAGAGGGACCGATGAGTACTACCAGTA
GTATTATGCTGTTGTTCATGATGCATAGGAATGGGGTTCGGAAGGAGTTGGTGGCATTAG
TGGCGTTGTTCACTCTGGTGGGATTGGAGCCATGGGCGTTGGTCGGATTGGTGTGTGTTG
GAAAGTATTCCGTGGATCAAAGGAGGAGGAAACGGATCGGTGGTATGCCCAAGAAGGTCA
AGGTTGTGGAATCGTATTATGCACATGGTGTGGTTGGGGAGGAGGAAGAGGAGAGCGAAG
AAGTGGAACGGAGCAAAAAGTATGCTATTTTGGAAAAGCCCGTGGGTACCATCTTCAATC
CGGCGGATTTAAGTTTGAGAGATGAAGAGTACGATGTCATCTTGTTGGGATGTGGTCCGG
AAGTGCTGTACACTGCATCGTTGCTATCACGAGCGGGCAAAAAGACGTTGGTATTGTCAC
CACGAGAGGATGCCTCTGGGTGCTTGACATTGCAGAACGGAAAGACGAATGTTCCTTTTG
ATATTGACGGAAGTAACATTGCCCATTTGGCAAGACAACAGTCGCTGTTGGCTCCGGCAC
TTTGTACAACGACGGATACTCAGGGTGGCATTCGCTTTGCACGCATCGGAAGCGAGGTGG
ATGGCTATGCTCATTCTATTCTCTCGGTGCCAGGCCTGGGCACGGATTCTATCTCGAATG
AGTGTATTCCGATCGTATTGACGGCGGAAGGTGAAGTTGCCTTGGCGGAGTACTGCTCGA
CGTATCTTGGAGATGCATTTCCTGGTACTGATTTGGATGGAAACGATAATGGCAACTCTA
CTTCATTGAGCTATCTCAAGGCATGTGGACAAATCAACGCTGGGTCTGGTGACTTTTACT
TGGCCAAGCTTTTTCCCAAGGCAGCTGAATCGTTCAAATCTTCCGACTCCAATGTCTACC
AACAAGCTTCTATTCGTCCAGCATCGACATTCCTTAACAAGTGCCTACCGCTCAATACTC
ACGTGCGTGCTCTCATGGCGGCCATTGGAATGGCGAATGAGAACTTGAGTCCAGATAAGA

CGAGCATGGCGGCTCATGTTACCAATGTTTGTGCCATGACTAGTACGGAGGGCTATGCGT
ATCCTGTTGGAGGTCCGAGGGCGTTGTGTCATGCATTGACGAGCGTGATAGAACAGAATG
GTGGAAGAGTTGTGAGTGGTGTTTTGTTGCAGGAGTTGCTGTTTGAGAAGTTGGAAAAGA
AGGAACCAAAGGAAGAGACTAAAGATGGCGAGTCAAAGGAGCCAAAGCCTCGTTGCAAGG
GGATAAGATTGGAGAATGGCTTAGAGTTGTCGGTTTCAGACAAGGGAGCTGTCGTTTCGT
TCATGGGAATGATACCCACCTTTTTGCAACTCGTATCTCCTGATGTACGAACTGCCGAGG
GAGTTCCTGCCGGCCTGCCAGCACTAGAGGAACGCCGTCCCTTGATGAGGGTCATGATTA
GTCTCAAAGGAAATAAGGACGACTTGAACTTGACGGGAGCCGATTGGTATCGCTTGCCCA
ATGCCACCTTGCCGAGGGATGAGTTGGATCCAATGACCGGTCAGGTAAAATTCGGAACGA
TTGGTGTAGACGACGATAATACTGGCGCAAGCGAGGAGTTGATACTTGGCGAAGCTACAG
ATGAGACAGAAGCAACGACTAGTCACACACGAGGCAAGCGAAACAAAGCAGCCACGTCGA
AGGCGCCACGATCCAAGTTTACATCTGGAGTATCATGGATGAAAGTATCATTTCCAAGTG
CCAAGGATCCAAGTTGGCAAGATCGACATGGAGACGTTTCCACTTGCGTTGTAACAGTCG
AGGCAGACGACGACTTTGTCCAAATGTTTGATACAAAGCCAAAGATTTACTCGGTGTTGA
AAGCGGGTAATGGCGAGAGAGAACGATTGCGGGACCGAGTGTTGAAAGATTTATTGGAGA
CATTTCCTCAGCTTCAAG

Exon 2:
>Thaps3 chr_18:364943-364515
GCCAGTTGGAGACTGTCCAGATATGTGGACCCGTGCGGTCTGGGCTTACTCACAATGGTC
CCAGATTTGCCATCAAAGGAAATCGTCCAGAAACTCCGTACCCCGGTCTGTACATTGGTG
GAGCGGATCTTACTGTGGGTGATTCCTTCTCTGGTGCAATCGTTGGTGGATGGTTGGCTG
CTAATGCGATCATGGGTTACAGTTTCATGGATCATATGTATCTCGGGAAGAACATCACTT
CGGACCTGCAGCAGTTCATAGAGGAACCGATTTTGGCAACTGAAAGGAATGGTGTCATAG
TGGATGACGTTGCTGTTCCTTTCAAGGAGGTTGTTGTTGATATGCAGAAAGGAATCACGG
ATGCAGATAGAAGCACCGCAGCTGAATCTAGTAAAGAGGAGTAATCGTAATCATAGGATG
CTATTGAAT

Translated:

MRMGRPNKKLRSTSKQTTNPNPPKYSSPTLVVGQVSSNVIHSIYGPALTKLAVESVEE
YADAVLRWEASLPEVLVKPSQLDDAEDIDVDADGTFEKGEEVEVDLDGSILPSHDNDDDK
TSSPTMPTSNRLFTTQSSIDNLTALLTDTSQHFSTTNAWKIHANAAKFERLLDEKYGRFR
PFIESHPELEVFIKKVQRKYAMGQFSPLRKGEGPMSTTSSIMLLFMMHRNGVRKELVALV
ALFTLVGLEPWALVGLVCVGKYSVDQRRRKRIGGMPKKVKVVESYYAHGVVGEEEEESEE
VERSKKYAILEKPVGTIFNPADLSLRDEEYDVILLGCGPEVLYTASLLSRAGKKTLVLSP
REDASGCLTLQNGKTNVPFDIDGSNIAHLARQQSLLAPALCTTTDTQGGIRFARIGSEVD
GYAHSILSVPGLGTDSISNECIPIVLTAEGEVALAEYCSTYLGDAFPGTDLDGNDNGNST
SLSYLKACGQINAGSGDFYLAKLFPKAAESFKSSDSNVYQQASIRPASTFLNKCLPLNTH
VRALMAAIGMANENLSPDKTSMAAHVTNVCAMTSTEGYAYPVGGPRALCHALTSVIEQNG
GRVVSGVLLQELLFEKLEKKEPKEETKDGESKEPKPRCKGIRLENGLELSVSDKGAVVSF
MGMIPTFLQLVSPDVRTAEGVPAGLPALEERRPLMRVMISLKGNKDDLNLTGADWYRLPN
ATLPRDELDPMTGQVKFGTIGVDDDNTGASEELILGEATDETEATTSHTRGKRNKAATSK
APRSKFTSGVSWMKVSFPSAKDPSWQDRHGDVSTCVVTVEADDDFVQMFDTKPKIYSVLK
AGNGERERLRDRVLKDLLETFPQLQGQLETVQICGPVRSGLTHNGPRFAIKGNRPETPYP

GLYIGGADLTVGDSFSGAIVGGWLAANAIMGYSFMDHMYLGKNITSDLQQFIEEPILATE
RNGVIVDDVAVPFKEVVVDMQKGITDADRSTAAESSKEE-

Targeting:

SignalP 3.0 - NN: No

SignalP 3.0 - HMM: No, Signal Peptide Probability = 0.004

SignalP 4.1: No, D = 0.101 (D-cutoff = 0.450)

ChloroP: Yes, Score = 0.543

Analysis using peptide sequences starting with any of the next four in-frame methionines downstream the first one did not result in clear predicted ER or chloroplast targeting.

19)Thaps3_14875

Most of the region was not covered by RNAseq reads. However, there was an open reading frame, that encompassed an approximately 450 base pair region with reads mapped to it.

Open reading frame:
>Thaps3 chr_3:2375339-2377698
ATGGACAAAAATAGAGATGCCTTGGCCGACGAGAACGCCGCACCGCGAAAGCATTTGGCA
CTCCATCTCACTCATAACCGTGTTCATCATCACTGGCACACCAACGATGGACAATAGACC
TCTCTCCTGTCTGACGCCAGAGTGCACAATAGATCGCCATCCCAACGCTACTACTTCTTC
GTTTCAATCGTTTGGAGGTGCAGCAACCTGCAGTGGATCCTCCACTGCATCCACGACGAC
TCATCTACTCCTGTCATCCATTCTCATCGGCCTCCTCTCTCCCATCATCTCATGTCTCTT
CGTCATTCTGCTCCTCATGTTCTTCAAACGACGTCAAACCCGTCATGAATTGGAAGGAGG
CAAGTGCAATCTACCCACTGTCGTTTGGAGACCAAGATTTATGAACTACACGTCCAAGGA
TGAAGGATCCGATGAGGACATTGAGATCGATGATTATGAAGCATGGGCAAGAGAGTATGC
ACGTTCACTGCAGTCAGACGACGGCAACAACAGCGGCACATCAACTCATATGAAGAAGTT
GGGATCCTCGGCAATAACCAACATATTACCAAGAATGGAACGCCTCAATGGTCCGTATGG
AATGTACGCCACCGTTTACGGAGTGTCCACGAAGGTATTGCACGTGGCTCATCCAGTTCC
CGCCAGAGCTATTCTGACGGGGAGTGGAGTTGTAGATGTTGGAGGAATGAACAATGGAAT
CGGAGAATGCTTTGAACGACAGAATAGTTCGTTTTTAGGAGAAATATCACAATCAGTGAC
TCGGCCGTTCAAACGATTGTCATCGGGAATGGAGGGTGCTGCTATCAGTCCATCCGAAGA
ACGAAAACAGCGTCGTAGATCTTCTGCACTGCGATTGCTCACCGGCTCAACAAAGTATCC
AGCCTATGACCACTTTAAGAACTTTTCAGGGGATGGAGTCTTCACCGCTGATGGTTCTGA
CTGGAAAGCGAAACGTGCTAGCGTCTTGCACTGTTTATTGAGAAGTGGGGGGGCCGATTG
CATGTTGGAAAAGGAGATTAATAGGGCTGCTGACTCTTTTGAGAGGGAGGTTACGTGGGC
GAAACAAACAATGAATAAGGAGGGCGATGATAAGGATGGTCCAGTGATGAATGTGGTGAC
AATGTTACAAAGGTCGACGATTGGTCTCATTTATCGCATCATTACACACCACAATGTGGA
GTTCAGTCCAGACATTGATACAAACGAGCAATTCATTTGTTCTCCAAAGAGCTCAGCAGC
ATCTCTCACATCCTTGGACAAGAACCAACACAACGGTGCTAAGGCATCAGAAGATGATAA
CCACACCAAACCCGATGTAAAGAAGGACTCACAGATGAAGTTACTTCTACCAATCTACCT

CGATGCAGTCACCAAAATACGAATGATTGTCCTCGCTCAGTCCAGATCTATTTGGTATCT
TCTGCCACGATGGGCCTATCGCACATTCTCTCCCATGTATCGTGACGAAGAAAGAACAAT
GGTTCCGATTAGACAGTTTGCCAGATTGGCGTGTGAGAATGCAGTGGAGGGAAGCCCGTT
GGAATTGCTGAGTCAAAGGAGTAGTCACGCTTCAAAAGAGGGCGAAGCGACCAGTGCAGT
CTCGAAGGATTTGTTGGATGAGGCCATTACTCTCCTATTTGCTGGACAGGATACTTCTGC
TGCCACCTTGTCGTGGACACTGCATCTACTCTCACTTCATCCACAGGAGCAGCAAAAGGT
AGTGGAGGAGGTTCGTTCAGTACTGTCATCTTTGGATGAGGGCGAAATGGTATCCAAGAA
CACCATCTCTCAGCTGCCATATTTGGATGCAGTCATCAAGGAATCGATGAGACTTTATCC
TGTTGCACCATTCATCGTTCGAAAGCTTACCACGGACATGACTATTCCCATCGAAAGTCA
GTCTGTAGAAGATGATGCCACAACAACTACCATCCCCGAATCAACCTTTGCATGCATATG
GATATACGCACTCCAACGAAACCCCAAGCTATGGACAAAACCAGACGAATTCATCCCCGA
ACGATGGATCGATCCTGATCTACGAAGCAACGACCTCGGCCAACAAGAGGTTGGCTCATA
CATGCCATTTGCGCTCGGTCCTCGTAATTGCTTGGGGCAACCAATAGCTCAAGTCATCTT
AAGAGTACTATTGGCGAGGATACTGAACAAGTATGAAGTGAGGGATCCCAAGTTTGATGC
CTTGCAGAGGTTGGGGGAGGAAACGGGGGAGGCATTTGATACCAAGTATCTTCTCAAGGA
TATGCAAGCAGGATTTACTGTTCTTCCTTCAAACGGATTGAGAATCAAGTTAGTGGAGAG
GTGCTAATTGTAGTGGGGTT

Translated:

MPWPTRTPHRESIWHSISLITVFIITGTPTMDNRPLSCLTPECTIDRHPNATTSS
FQSFGGAATCSGSSTASTTTHLLLSSILIGLLSPIISCLFVILLLMFFKRRQTRHELEGG
KCNLPTVVWRPRFMNYTSKDEGSDEDIEIDDYEAWAREYARSLQSDDGNNSGTSTHMKKL
GSSAITNILPRMERLNGPYGMYATVYGVSTKVLHVAHPVPARAILTGSGVVDVGGMNNGI
GECFERQNSSFLGEISQSVTRPFKRLSSGMEGAAISPSEERKQRRRSSALRLLTGSTKYP
AYDHFKNFSGDGVFTADGSDWKAKRASVLHCLLRSGGADCMLEKEINRAADSFEREVTWA
KQTMNKEGDDKDGPVMNVVTMLQRSTIGLIYRIITHHNVEFSPDIDTNEQFICSPKSSAA
SLTSLDKNQHNGAKASEDDNHTKPDVKKDSQMKLLLPIYLDAVTKIRMIVLAQSRSIWYL
LPRWAYRTFSPMYRDEERTMVPIRQFARLACENAVEGSPLELLSQRSSHASKEGEATSAV
SKDLLDEAITLLFAGQDTSAATLSWTLHLLSLHPQEQQKVVEEVRSVLSSLDEGEMVSKN
TISQLPYLDAVIKESMRLYPVAPFIVRKLTTDMTIPIESQSVEDDATTTTIPESTFACIW
IYALQRNPKLWTKPDEFIPERWIDPDLRSNDLGQQEVGSYMPFALGPRNCLGQPIAQVIL
RVLLARILNKYEVRDPKFDALQRLGEETGEAFDTKYLLKDMQAGFTVLPSNGLRIKLVER
C-

Targeting:

SignalP 3.0 - NN: Yes, Cleavage Site TPT-MD

SignalP 3.0 - HMM: Yes, Cleavage Site TPT-MD, Signal Peptide Probability = 0.147, Signal Anchor Probability = 0.762

SignalP 4.1: Yes, Cleavage Site ITG-TP, D = 0.475 (D-cutoff = 0.450)

ChloroP: No, Score = 0.473

Analysis using peptide sequences starting with any of the next four in-frame methionines downstream the first one did not result in clear predicted ER and chloroplast targeting.

20)Thaps3_bd_518

No available RNA-seq data for the unmapped "bottom drawer" sequences.

Gene model predicted by JGI (below) cuts off abruptly on the C-terminal side due to the way the unmapped sequences were assembled. Some C-terminal sequence is likely missing, as evidenced by the lack of an in-frame stop codon. The translated product of the JGI model does not start with a methionine, but can be extended to one upstream. There is only one such methionine that is in frame after the closest stop codon.

JGI Model:
>Thaps3 bd_35x67:17227-17611

AAATGCTATTGGCATTGACTGGGGACTCTCTTTGCAATACAACTT
ACCAATGCTTGCAAAAGTGTGATTACCTTGATGCAGTGGCTCGTGAAACGCTACGCCTTTATCCTCCGGC
TGCAAGCACTCGTTGGGCGACAGATGCAAAGGGTGCGAATGCAGGTGGCTTCAACTTGGAAAAGAGTGTT
GTTCATGTCAACTTCTATGCAATTCAGCGAGATCCTGACGTTTGGGAGAATCCCGTCTCGTTTGTTCCTG
AACGTTTCCTTGGCGAAGAAGGAAGGAAGAGGATACTGTCGTATTCGTTCTTGCCATTCAGTAAAGGATC
ACGCGACTGCATTGGCAAGT

Translated:

MLLALTGDSLCNTTYQCLQKCDYLDAVARETLRLYPPAASTRWATDAKGANAGGFNLEKS
VVHVNFYAIQRDPDVWENPVSFVPERFLGEEGRKRILSYSFLPFSKGSRDCIGK
SignalP 3.0 - NN: =No

SignalP 3.0 - HMM: No, Signal Peptide Probability = 0.004

SignalP 4.1: No, D = 0.203 (D-cutoff = 0.450)

ChloroP: No, Score = 0.440

If translated in a different frame, starting downstream the start codon used above, the peptide fragment is not targeted to the ER, and therefore, the chloroplast.

MQWLVKRYAFILRLQALVGRQMQRVRMQVASTWKR
VLFMSTSMQFSEILTFGRIPSRLFLNVSLAKKEGRGYCRIRSCHSVKDHATALAS

Targeting:

SignalP 3.0 - NN: Yes, Cleavage Site LVG-RQ

SignalP 3.0 - HMM: No, Signal Peptide Probability = 0.070

SignalP 4.1: No, D = 0.381 (D-cutoff = 0.450)

ChloroP: Yes, Score = 0.531

21)Thaps3_6395

No introns.

>Thaps3 chr_6:1467629-1469626
GCTTGATGAGGCTACCATTGGCATTGCTTTCAGCTTGTACTTCGGCTTCATTTCGGAATC
GTCATATCGTGCATAGCTTTACTTTCGGCAGGTCACATAGAGATTGCACAGGAATAATCC
CATCGTTTCGCTTTCGGGTGTCTCCTACAATGTCCTCCTCTGATAGACACAGCTCACAAC
TGAATCAACCCAAACGACCTCGTCACGAGGAACCATCACACAGTATCATGGCATTTGACT
TGAAGAATAGATCTCCGTACGAGAAGAAAATGGAGAGAGTCATTACAGCGTGTCCAAAGT
GTAACGGAGAAGGAAAAGTGCGAGCTCCGTTATCAAAGAAGGCTCGTGCCCAACGCAAAC
GAATGCAACAGAGCCAAACAGGAGATACAACTAATGCACCAAATCTGGCTATTCTGAAGA
AACCGTGTAAGGAGTGTGATGGATCTGGTTTGATTGCCATCAATCCTTTGGATACAACCG
AAAGAAAGCAGACACCACCACAGATTCAACCCAACTTTTCGGTAGCCATTGTAGGCGGTG
GTATTGGTGGCATTGCATTGGCTGCCGCACTACAGCATCGCAACATTCCATGTATTGTTT
ATGAACGAGATTTGTCGTTTGAGGAAAGAAAACAGGGATACGGACTAACGATGCAACAAG
GAGCACGAGCTCTAAGATCCTTGGGCTTCTTTTCATTCTCTGACGATGGAGAGGACGACA
ACAACAATTGTAGTGGCAAAAAAGCAGTGGATGAGAATACTTCAAATACAAAGCAAAAGT
TTGGAATCCACTCAACTCGTCACGTAGTTCACAAGCCAGATGGAACTGTAGTAGGTGAAT
GGGGTATGAAAGTCTGGGGTGGTCGATTCGAGAAGAACGGCAGGAAGCACGCCAAGCGAC
AAAATGCACACATCTCTAGGCAAAATCTTCGCCAGCTGTTGATGGAGATGCTGCATCCTG
GTACAATACAATGGGGGCAAAAGTTTGTGGGTTATTCGGGACAGTCTAGTGATGACGATT
CCTCACAGGATCAACCATCATTGCAAGTCAGATTTCGACGCAGAAGTAACGATTGTGATG
AGGAGGTCGCTACAACTGCGTCTGTACTTGTCGGATGCGATGGTATCCGATCGTCTGTAC
GATCTGCAAAGTTGGGTGAGGACGGAACACCACTCCGTTACCTGGATTGCATTGTCATTC
TTGGCATTGCTCCGTCGCCAACCTCGGCGTTAACTGATGGTGAAACTGTATTTCAGACGG
CAGATGGCATAACTCGTCTGTATGTCATGCCGTTTGCGGAAGCTGGAGATGACTCGTCTG
GTTTATCAACTGACAACACTAAAGGATTGAGCATGTGGCAGCTCTCGTTTCCGATGGACG
AGACTGATGCAACAAGGCTGAGTCAACTTGGATCGTCTGCGTTGAAAGAAGAGGCCCTCA
AACGATGTGGTGCATGGCATGATCCAATATTAAAGCTGTTACGTTCCACACCAGAGGATT
TCATTACTGGATATCCGTGTTATGATCGTGCCCTTGTCGAGAGAAAAGAGCTTCGAGATG
GATGTGATAAATCTCAATCTGCAAACGCCTTTGTGACTCTACTCGGCGATGCTTGTCATC
CTATGTCCCCCTTCAAAGGTCAAGGGGCGAATCAAGCTCTTTTGGATGCCGTGCTATTAA
GCCAAAAGCTCTTCGATATATCTCGTATTCATAACGGGAAAACGAACGTCAACGAACAGC
AACCTACCATATCACTCAATGAAAGCACACCACAGGCATTGGCAGAGTTCGAAAACGACA
TGCTACAAAGGTGTGAAGTCAAAGTTAAAAAGTCGGCAGATGCAGCAAAGTTCTTGCATA
GTGACGTCGCTATTCAAGAGGGAAACATTACACGAGGAGCGGCAGCGTTGGATGCGAAGG
GATGAGAGCGAGATTCTTCTGGGATCTAGTTGGATACAATCATAAGATAACCTAAGTTAC
CCCTGTAATGAGCGAACA

Translated:

MRLPLALLSACTSASFRNRHIVHSFTFGRSHRDCTGIIPSFRFRVSPTMSSSDRHSSQL
NQPKRPRHEEPSHSIMAFDLKNRSPYEKKMERVITACPKCNGEGKVRAPLSKKARAQRKR
MQQSQTGDTTNAPNLAILKKPCKECDGSGLIAINPLDTTERKQTPPQIQPNFSVAIVGGG
IGGIALAAALQHRNIPCIVYERDLSFEERKQGYGLTMQQGARALRSLGFFSFSDDGEDDN
NNCSGKKAVDENTSNTKQKFGIHSTRHVVHKPDGTVVGEWGMKVWGGRFEKNGRKHAKRQ
NAHISRQNLRQLLMEMLHPGTIQWGQKFVGYSGQSSDDDSSQDQPSLQVRFRRRSNDCDE
EVATTASVLVGCDGIRSSVRSAKLGEDGTPLRYLDCIVILGIAPSPTSALTDGETVFQTA
DGITRLYVMPFAEAGDDSSGLSTDNTKGLSMWQLSFPMDETDATRLSQLGSSALKEEALK
RCGAWHDPILKLLRSTPEDFITGYPCYDRALVERKELRDGCDKSQSANAFVTLLGDACHP
MSPFKGQGANQALLDAVLLSQKLFDISRIHNGKTNVNEQQPTISLNESTPQALAEFENDM
LQRCEVKVKKSADAAKFLHSDVAIQEGNITRGAAALDAKG-

Targeting:

SignalP 3.0 - NN: No

SignalP 3.0 - HMM: Yes, Cleavage Site VHS-FT, Signal Peptide Probability = 0.974

SignalP 4.1: No, D = 0.207 (D-cutoff = 0.450)

ChloroP: Yes, Score = 0.554

Analysis using peptide sequences starting with either of the next two in-frame methionines downstream the first one did not result in a more clear targeting prediction.

22)Thaps3_10233

Exon 1:
>Thaps3 chr_15:679062-677604
CGTTGACTCTCTTTCAACGACGACACCCACCAAGTCATCCATCTTAAACTTGTTGGTAGT
ACGCCACTCTTGTTGAAAAGTAGATCCAACGACCATGATCGGACGCAAGTACAGTCTCGC
AGCCTCAGCTCTGGCAATTATCGCCTCCCTCACCTCCACAACCGCATTTGCACCACCCTC
CTCCTCTCTCCTACGCTCCACTCGCCTCCACTCCACCGTCGAAGAAACCACCAACGGAGA
AGCCGCAACAAACACCAATGTCGAACAAATCAAAGACACCTCACGTGACAAAGTAATGAC
ATTTTCCTACGACATGTCCATTGAACCCAAGTACGAAAAGCCAACGTATCCAGGAACAGG
CAACGGCATGTCAGGCGACTCTGGTGAATACGACATTATCGTTATTGGATCAGGAATGGG
CGGACTCGCCTGTTCGGCACTCTCTGCCAAGTACGGCTCACGAGTCTTATGTTTGGAAAG
TCACATTAAAGTGGGAGGAAGTGCTCATACCTTTAGCCGAATGCACAATGGGGGCAAGTA
CAGTTTCGAAGTCGGACCGAGTATCTTTGAGGGATTGGATCGTCCATCGTTGAATCCACT
GAGAATGATCTTTGACATTTTGGAGGAAACCATGCCGGTCAAAACGTACAAGGGATTGGG
ATATTGGACTCCTTCTGGCTACTGGCGTTTCCCCATTGGATCACGTGAGGGATTTGAACA
GTTGTTGATGGAACAGTGTGGTGAGGATGGAGAAAAGGCTATTGGAGAGTGGAAGGCGTT
GAGAGAGCGATTGAGGACTTTGGGTGGAAGTACGCAGGCCGTGGCATTGTTGAATTTGAG
GCAAGATGCGGGATTTTTGGCAACTACTGCCGGTTCATTGCCGTTTGTGGTGACTCATCC

TGATGTGTTTGGGGATTTGTCATTGACGTTTGATGATTTGAGCAAGACGGTCGATGAGTT
TGTGACTGTGCCTTTCTTGAGGAATTTTATTGATACGATGTGCATCTTTTGTGGATTCCC
TGCAAAGGGAGCCATGACGGCACACTTGTTGTACATTCTCGAGAGGTTCTTTGAGGAGAC
TGCAGCTTTCTCTGTTCCTATTGGTGGAACATGCGAGCTTGGAAACACTCTTCAGCGTGG
ATTGGAGAAGTATGGAGGTAAATTGCAACTCAATGCTCATGTCGATGAAATTCTCGTAGA
GAACGGACGTGCCGTCGGAGTCCGTCTCATGAATGGAAATGTCGTCAAGGCACGCAAGGC
AGTGGTCAGTAATGCTACTCCCTTTGATACTGTCAAGTTGATGCCCAAGGCGGAGGGTGA
GCCCAAGGGATTGACTAAGTGGAGAGAGGAGTTGGGAAAGCTTCCTAGGCATGGAGCTAT
CAGTCATTTGTTCTTGGCGATTGATGCGGAGGGCTTGGACTTGAGTCATATTCAGGATCC
TGCTCATTTGGTAGTTCAG

Exon 2:
>Thaps3 chr_15:677527-676741
GATTGGGATCGTTCTCTTCAAGACTCCCAAAATCTTTGCTCCTTCTTCATCCCATCCATC
CTCGACAAAACGCTATGCCCAGAAGGCAAGCACGTCATTCATGTCTACTCTTCTGGTGGA
GAACCTTACGAACCATGGGAAAAGCTCACACCTGGCTCAGAAGAATACGAAGCATACAAG
AACGAACGTGCCGAGGTTCTCTGGCGTGCTGTGGAGCGTTGCATTCCCGATGTTCGTGAT
CGTGTGGAGTTTTCTATCGTTGGTTCACCTTTGGCTCACGAGGCTTTCCTCCGCAGAGAT
AGGGGAACATACGGTATGGCTTGGGCTGCTGGTTCGTCGGCTCCTCAATCTGGTATCCTT
GGAAGTGTTCTTCCTTTCCCATTCCCCAACTTGAAGACTCCAGTGGACGGTTTGTTGCGT
TGTGGTGACTCATGCTTCCCTGGTATTGGAACTCCGTCGGCTGCTGCTAGTGGTGCCATT
GCTGCCAATACAATGACTCACGTTGATAACCATTTGAAGATGCTGTCGGAGGCGAGTAAG
TTGGATCCTATGTACAAATTCTTGGATGCTGGTATCATGGGGCAGGTTTACAAACCCTTG
GTGCAGGGATTCACTCCTAGTCCAGAGTTGAGGACCGACCAATACGTTTCGGGTGCTGGG
GTTGCACCTGTCGATTACACTGCCACTGATCCTAGTGTCAGCGAGAGGATTGATTTGTAA
GAGTGGTAAAATAGTGGGGGTAAATATAGCGTAGTCAGAGTTATTTGTCACGTGAAGAGA
GCAATTG

Translated:

MIGRKYSLAASALAIIASLTSTTAFAPPS
SSLLRSTRLHSTVEETTNGEAATNTNVEQIKDTSRDKVMTFSYDMSIEPKYEKPTYPGTG
NGMSGDSGEYDIIVIGSGMGGLACSALSAKYGSRVLCLESHIKVGGSAHTFSRMHNGGKY
SFEVGPSIFEGLDRPSLNPLRMIFDILEETMPVKTYKGLGYWTPSGYWRFPIGSREGFEQ
LLMEQCGEDGEKAIGEWKALRERLRTLGGSTQAVALLNLRQDAGFLATTAGSLPFVVTHP
DVFGDLSLTFDDLSKTVDEFVTVPFLRNFIDTMCIFCGFPAKGAMTAHLLYILERFFEET
AAFSVPIGGTCELGNTLQRGLEKYGGKLQLNAHVDEILVENGRAVGVRLMNGNVVKARKA
VVSNATPFDTVKLMPKAEGEPKGLTKWREELGKLPRHGAISHLFLAIDAEGLDLSHIQDP
AHLVVQDWDRSLQDSQNLCSFFIPSILDKTLCPEGKHVIHVYSSGGEPYEPWEKLTPGSE
EYEAYKNERAEVLWRAVERCIPDVRDRVEFSIVGSPLAHEAFLRRDRGTYGMAWAAGSSA
PQSGILGSVLPFPFPNLKTPVDGLLRCGDSCFPGIGTPSAAASGAIAANTMTHVDNHLKM
LSEASKLDPMYKFLDAGIMGQVYKPLVQGFTPSPELRTDQYVSGAGVAPVDYTATDPSVS
ERIDL-

Targeting:

SignalP 3.0 - NN: Yes, Cleavage Site TTA-FA

SignalP 3.0 - HMM: Yes, Cleavage Site TTA-FA, Signal Peptide Probability = 0.998

SignalP 4.1: No, D = 0.619 (D-cutoff = 0.450)

ChloroP: Yes, Score = 0.560

23)Thaps3_33926

Exon 1:
>Thaps3 chr_4:1242467-1242109
TTCGCATTTTGCTCCTCGTCCTCCTCCTCACAGCTCTCATCCATCCAACACCACACCACA
CTACCGCAATGACAATCCCCGGCACCACCACAAGCGGTATGCCCTTCGGCTTCGGCTCCT
CCACAGGACCATCCGACAACCTCCTAATCGGCCTAACATGCCTCGCCATCTTCTCCCTCT
TTTACGCCTTCTTCGTCGTCCGCCCTAGGACGAAGGGAGGACCACATGCACCACCGGTGG
TTACGTCGAGCCCTGTCTCAAGTCTTCCTGTCGTTGGGACTATTGTGGAGTTTGGCAAGA
GTCCTGTGAAGATGGTTCAGAGGTGTTATGAGGATTACGGTCCTGTGTTTACTGTGCCG

Exon 2:
>Thaps3 chr_4:1241947-1240596
TTCTTCCACAAACGTCTCACCTTCCTCATCGGCCCCGAAGCCCAAGAACCATTCTTCAAA
GCACCCGACGAAGTCCTCTCCCAAAACGAAGTCTACGGCTTTATGAAACCCGTCTTTGGA
CCTGGAATCGTCTACGATGCCTCCAAAAAGAACCGTCAAGTTCAATTCCAATCAATGGCT
AACGGCCTTCGCACTGCTCGTCTTAAGGGATACACTGCCAAGATTGAACGTGAGACGCGT
CAGTACCTCGAATCTTGGGGAGAGTCCGGAGAGCTTGATCTATTCCATGCTCTTTCGGAG
TTGACTATTCTTACTGCCTCTCGTTGTCTTCACGGAGATGATGTTCGTGAAAATCTCTTC
AAGGAGGTTTCGGAATTGTACCACGATCTTGACCAGGGCTTGACTCCACTCACCGTGTTC
TTCCCCAATGCTCCTACAAAGTCTCACATGAAACGCAATGCGGCACGTGCCAAAATGGTG
GAGTTGTTCTCCAAAGTGATTAAGAATCGTAGGGATAATCCTGATGTGCAACACTCGGAT
GGTACGGATATCCTCTCCATCTTCATGGATGTCAAGTACAAAGATGGATCAAACATTACC
GACGAGCAAGTGACTGGGCTTTTGATCGCATTGTTGTTCGCTGGGCAGCATACGAGTTGC
ATTACTTCTACATGGACGAGTCTCTTTATACTCAACAACCCTGCCATTCTCAAGCGTATT
ATTGCTGAGCAGAATGACGTCTTTGGTTCTCAACCGGATGCCGATGTGGATTACAAGATG
GTGAACGAGGATATGCCCTTGTTGCACAACTCGATGAAGGAGGCTTTGCGTTTGTGCCCT
CCGTTGATTCTTCTCATCCGTTATGCTCTCAAAGACGTGAAGGTGAAAGCTGCCGGAAAG
GACTACACCATTCCTAAGGGCGATATGGTGCTCATTAGTCCATCTGTTGGTATGAGGATT
CCCGAGGTGTTTAAGGAACCTAATACCTTTGATCCTGATCGTTTCGGTCCTGATAGGGAG
GAGGACAAGTCAAGTCCATTCGCTTACATGGGCTTTGGAGGAGGTATGCACAGTTGCATG
GGACAGAACTTTGCGTTTGTTCAGGTCAAGACGATTCTTAGCGTGTTGTTCCGTGAGTTT
GAGTTGGAGATGGTTTCGGAGACGATGCCCGACATTGATTATGAAGCCATGGTTGTTGGA
CCCAAGGGAGATTGCCGTGTTAGGTACAAGAGGCGTCAGTAGATGATGGGTACCATGTCG
AGTTGCAACTGTCTTGATCAAACATAGATGTTTTGAGAAACTGGGGTTGATTTCATCGTA
TTGATTTTAATTGAATAAAGAAGCTAGAGATT

Translated:

MTIPGTTTSGMPFGFGSSTGPSDNLLIGLTCLAIFSLF
YAFFVVRPRTKGGPHAPPVVTSSPVSSLPVVGTIVEFGKSPVKMVQRCYEDYGPVFTVPF
FHKRLTFLIGPEAQEPFFKAPDEVLSQNEVYGFMKPVFGPGIVYDASKKNRQVQFQSMAN
GLRTARLKGYTAKIERETRQYLESWGESGELDLFHALSELTILTASRCLHGDDVRENLFK
EVSELYHDLDQGLTPLTVFFPNAPTKSHMKRNAARAKMVELFSKVIKNRRDNPDVQHSDG
TDILSIFMDVKYKDGSNITDEQVTGLLIALLFAGQHTSCITSTWTSLFILNNPAILKRII
AEQNDVFGSQPDADVDYKMVNEDMPLLHNSMKEALRLCPPLILLIRYALKDVKVKAAGKD
YTIPKGDMVLISPSVGMRIPEVFKEPNTFDPDRFGPDREEDKSSPFAYMGFGGGMHSCMG
QNFAFVQVKTILSVLFREFELEMVSETMPDIDYEAMVVGPKGDCRVRYKRRQ-

Targeting:

SignalP 3.0 - NN: No

SignalP 3.0 - HMM: Signal Anchor Predicted, probability = 0.936

SignalP 4.1: No, D = 0.136 (D-cutoff = 0.500)

ChloroP: Yes, Score = 0.543

Analysis using peptide sequences starting with either of the next two in-frame methionines downstream the first one did not result in a more clear targeting prediction.

24)Thaps3_1549

No introns.

>Thaps3 chr_1:1808544-1810261
CTCGTCGACATCGAAAGACGGCAAACAGAAACAGCAACACAACGGAACAGCAGATTGATA
CATGGTGCATGTTTGATACCATGACAGCGCCATCGTCACCCTCCTCCTCTCGCTCTGCCC
TCGTCACGGCGGCAGCAGCATCCCTCGTCTCTCTCTCCTTGCTCTCCATCTACAAGCGTC
GCCGTACTACGTCCAACAATGAACTTCCCTACCCACCCACCCCTCCCGACAGGAACTACT
TCCTCGGCCATGCAATGTCTCTCCGACGAGTTCCCGGAGAACCAAAGAAATCGCACGATC
TTCTCTTCTTGAACTGGATGAACAAACTCAACAGCAAAGTAGTAATGTTTGAGCTCCCAT
TTCTTGGACGGCTGTTCGGTCTCGGGCGTATGATTTGCGTAGGTGATGCCGAGATTGCTC
GGCACATACTCGTTACGGCGAACTACAACAAGTCTCCCACCTACAGTGTGTTACAGCCAC
TCATCGGCATGAGTTCCATGGTTGCCACGGAGGGAAAGATGTGGAAGGATCAAAGAAAGT
TGTACAATCCTGGATTTTCTCCAGAGTTTCTTCGCAATTGTGTATCGACAATTATTGAGA
AGTGTAATAGATTCATCGCCAGATGTGATGGTGATGTTGAGAATGGTGTAGCGACGGATA
TGTTGGCGAGATCCATTGACCTCACTTCTGATGTGATTGTGCAGGTAGCATTTGGAGAGG
ACTGGGGAGTTGATAGCAAGGACAAACATGGTATCGAGACACTGCAAACAATACGAGATC
TTACGGTAGCCGTTGGGGAAAATATGACCAACCCATTACGCAAATACTTTGGATTACGAA
GCATTTGGAGAACGAGGAGACTCTCGGCAGCTCTGGATCAGGATATGCAAAATCTTGTGA
AGAGAAGACTTGCTCAGGTGTTGGCTGGAGATGCTGATTTAGAGAAGGATATCTTATCAT

TGACGTTGTCTGGTGTTTTGGAAGCAAAACAGGAATCTAAGTCAGGCGCCATCTCGTTGA
GCAAGGACGAAATGGAAAGGATGACATCGCAACTTAAGACTTTTTACTTCGCTGGGCACG
ATACCTCTTCATCCGCTATCGCATGGGCGTATTGGCTGTTGACAAAACATCCAGAATCAC
TCCAACGAGCTAGAGAGGAAGTTGTATCACACCTCGGAAGAGATTGGTCGGACGAGGCAT
TGACTGGGGACTCTCTTTGCAATACAACTTACCAATGCTTGCAAAAGTGTGAGTACCTTG
ATGCAGTGGCTCGTGAAACGCTACGCCTTTATCCTCCGGCTGCAAGCACTCGTTGGGCGA
CAGATGCAAAGGGTGCGAATGCAGGTGGCTTCAACTTGGAAAAGAGTGTTGTTCATGTCA
ACTTCTATGCAATTCAGCGAGATCCTGACGTTTGGGAGAATCCCGACTCGTTTGTTCCTG
AACGTTTCCTTGGCGAAGAAGGAAGGAAGAGGATACTGTCGTATTCGTTCTTGCCATTCA
GTAAAGGATCACGCGACTGCATTGGCAAGTACTTTGCTCTTCTTGAAATAAAGATTGCAT
TGGCTGCTTTGATTTCTCGGTACGATGCATCAGTTGTGAATGAAAATGAGCAGTATGTTA
TCCGTTTAACATCTGTTCCTCACGACGGATGCAAAGTGAATCTCTCTCGTCGCAGGAAAT
AAACGTGCTGTAAAGAAATCCTAAACTAACTTAGTATT

Translated:

MFDTMTAPSSPSSSRSALVTAAAASLVSLSLLSIYKRR
RTTSNNELPYPPTPPDRNYFLGHA MSLRRVPGEPKKSHDLLFLNWMNKLNSKVVMFELPF
LGRLFGLGRMICVGDAEIARHILVTANYNKSPTYSVLQPLIGMSSMVATEGKMWKDQRKL
YNPGFSPEFLRNCVSTIIEKCNRFIARCDGDVENGVATDMLARSIDLTSDVIVQVAFGED
WGVDSKDKHGIETLQTIRDLTVAVGENMTNPLRKYFGLRSIWRTRRLSAALDQDMQNLVK
RRLAQVLAGDADLEKDILSLTLSGVLEAKQESKSGAISLSKDEMERMTSQLKTFYFAGHD
TSSSAIAWAYWLLTKHPESLQRAREEVVSHLGRDWSDEALTGDSLCNTTYQCLQKCEYLD
AVARETLRLYPPAASTRWATDAKGANAGGFNLEKSVVHVNFYAIQRDPDVWENPDSFVPE
RFLGEEGRKRILSYSFLPFSKGSRDCIGKYFALLEIKIALAALISRYDASVVNENEQYVI
RLTSVPHDGCKVNLSRRRK-

Targeting:

SignalP 3.0 - NN: No

SignalP 3.0 - HMM: Signal Anchor Predicted, probability = 0.936

SignalP 4.1: No, D = 0.136 (D-cutoff = 0.500)

ChloroP: Yes, Score = 0.543

Analysis using peptide sequences starting with either of the next two in-frame methionines downstream the first one did not result in a more clear targeting prediction.

25)Thaps3_264647

Exon 1:
>Thaps3 chr_19c_29:130307-130912
CAAACATCACTATTTTTGGCAAGGGAGGATGTCTTCATTCTGAGTAGTCTCCAATATATT
TCGCTACAAATCCGGTCATACTCTCCTTATAATTTTTGATATGCCTTCTTCAAGGTTCGA

TGCTGCTCGATTTGTATCTTTGGTTGTTCACACAGTCAGCAGTGAGAAGTGAGGAGCTCT
CTCTGCATTTGGATATATAAAACTGATAGAAGAGATCAGCCAACACACATAACAATTTGG
TGGAGGAGCGAGAGATTTCATTGCACCATCGAGAGAGCGGCGAGGCAGGAGGCGACGGTG
CATCCATACTTGATCCATCGCATCGGACGGTCAGATTTGCAAATCGGCAATCGTACCGAC
ACATCAACTGCTCCCACTTCCCAGTGTTCCTTTCATCCCTGATCAGTCCTTCTTCCGCAG
CAGCGATAGTATGATCAGTAATATGACGATGGCCCTTTCAACACTGGCATTGTATCTTAC
ACCAACCACAATTATCACCCTTCTGCTATGTCTCTTAGTGACTCGGTTCATCCAATGGAA
GAATCACCTTGCGAATATGAAGAGCCAGACTCCCTTTCCTGGAATTCCAGTAGTTCCCGA
TGCTCA

Exon 2:
>Thaps3 chr_19c_29:131012-133019
CTGGCTACTCGGCCACTATCCCTTGTTCGTCAACCCCGACAAACACCACCAAACCTTCAC
CGCCCACGCCACCCCCTCCGGCATATCCGCCCTCTGGGGTCCGTCCACTGATAAATTCTT
CTCCTCCGTCCGAGCCGATCACTGCCGTTCCATCCTTCGTCAAAGCTCCAGTAGAAACTT
TGTCTCGTTCATTGTACGGCATGGACGGAGGACATTGGGAGAGGAGAGTATCATTTTGAT
TAACGGAGGAAAGAGGTGGAAGAGGCAACGAAAGGTGATTCAGAAGGCGTTTCATTTGGA
GGTTGTGAAAGGTAGGAGGGAGGCTGTGGGGGAGGTGGCGGATGTTGTTGTGGATTGGAT
ACTGAGGGCTTGTAGTGGTAGGAGTGATAGCGGTGTGCATGATGGAGATTGTGGAGGGAG
ATTGGTTGGTGCGGATGGGAATGAGAAGGTTTGTGTGGAGGCTGAAGACTTCTTTAAGTT
GTTTGCGTTGGAGGTGTTCGGAAAGGTAGCAATGGGATATGATTTTCGGTGCTTTCCTTC
TCTTGCTACTTCGGACGACAACGGCAACAGCAACAGTAATGTTGCCTTACACAACAAACA
AGATGTGGTGAGCAATGGCACCCATACATACAACGACAACGCGTGCAACTGTCTCCAAAT
GCCACCCGATGCACAGTCCTTTGACTTTCTGAATGTCGACATCGGGAACCGCTCAACACC
AACGAGCCTGATGAATCCGTGCATGCAATTCTACTCTATACCCACTCCACACAACAAAAA
GTATCATCATCATATGGACAGAATTAAGGGACTGGTTGGTAAGATTATCGGACTGCAACT
GAACAGGTTGTGCAGTGACGGCGGTGTCATGGAAGGCGATACGAACATGATTACCCACTT
GTTACAATCTACAATCGAAGAGAACTTCAGTCCTACCGATACAGATGGGAACACTGGCTG
TCCATTCTCATCATCCTTACCATCAAAGTCTATTCCAGATACAGTCATCTCCAACCTCAC
TCCATCAGATAAAGATCAAATCATTGAAAGCGTCTCCAAGATGCTCATCACATTCCTCAT
GGCAGGCTACGAAACCACTGCAATCTCAATGTCGTTTGTAGTGTACTTCCTTTCCAAATA
CAAACGATGCCAAGAAAGATGTGCAGAGGAAGCGAGAAGAGTTCTGGGGCGCTGTGGAGT
GCATGGAACGGACATTGACGATGACGAGCTGGTATACTGCCGTGCTGTCTTCATGGAAAC
GATACGGTTGCATCTGCCAGTCATGTTCACAACTCGTGTGACAGAGAAGGAAATGTCCTT
TGATACAGGGCTGGAGGAAGGCCATAATGTGACAATACCAAAGGGAACGAGGTGTGTCGT
TTGTCCTACAGTAGTTCATATGGATGAGCGTAATTTTGAACGAGCCGAGGAGTTCTTACC
TGAGAGATGGGTGCGGTGGGAGAGGGGCAGGTGGGTTGAACGAGATTACGAAACTGAAGG
ATTGAAGTCAACAGCATTGCCATCAATTACTGAAGATGAACAAGATTCTCCTCCCATATC
TGCAAAGTACGACGAAGAGAATAATTCTGCCAGTTCAATCTCTGCGGCTGATCCACACAA
TTTCTTCTCGTTCTCAGATGGAGCAAGGAATTGTGTTGGGAAACGTCTTGCAATTATGGA
GTCTACAATCTTGATTGCGGTATTGCTTCGTGACGTGTGCGTCGACTTCGCAGAGGAGGG
ATTTGAGATGAAAAAGGTACGGCGGTTCGTCACGTGTGGCCCTGAGAGTTTACCCGTCGT
GTTTTGGAGGAGGGAGTGAAGAAAGCTTGTAGTAGGTCTTAGTTTTTTTGCAGAAGAACC
GTAGCGAAATGCGGCTGTAACCAAGCTTCTTTAAGTAGACACACTATTGTGTGTAAAGTC

GTGATCATGTATTCCATGAAGCCGATCTAATGACTGGAGATGCTGTTCAGTTAAGATGCA
CTTACATGTGAAACTAACAGTTGAGGAC

Translated:

MISNMTMALSTLALYLTPTTIITLLLCLLVTRFIQWK
NHLAN MKSQTPFPGIPVVPDAHWLLGHYPLFVNPDKHHQTFTAHATPSGISALWGPSTDK
FFSSVRADHCRSILRQSSSRNFVSFIVRHGRRTLGEESIILINGGKRWKRQRKVIQKAFH
LEVVKGRREAVGEVADVVVDWILRACSGRSDSGVHDGDCGGRLVGADGNEKVCVEAEDFF
KLFALEVFGKVAMGYDFRCFPSLATSDDNGNSNSNVALHNKQDVVSNGTHTYNDNACNCL
QMPPDAQSFDFLNVDIGNRSTPTSLMNPCMQFYSIPTPHNKKYHHHMDRIKGLVGKIIGL
QLNRLCSDGGVMEGDTNMITHLLQSTIEENFSPTDTDGNTGCPFSSSLPSKSIPDTVISN
LTPSDKDQIIESVSKMLITFLMAGYETTAISMSFVVYFLSKYKRCQERCAEEARRVLGRC
GVHGTDIDDDELVYCRAVFMETIRLHLPVMFTTRVTEKEMSFDTGLEEGHNVTIPKGTRC
VVCPTVVHMDERNFERAEEFLPERWVRWERGRWVERDYETEGLKSTALPSITEDEQDSPP
ISAKYDEENNSASSISAADPHNFFSFSDGARNCVGKRLAIMESTILIAVLLRDVCVDFAE
EGFEMKKVRRFVTCGPESLPVVFWRRE-

Targeting:

SignalP 3.0 - NN: Unclear

SignalP 3.0 - HMM: Signal Anchor Predicted, probability = 0.894

SignalP 4.1: No, D = 0.340 (D-cutoff = 0.500)

ChloroP: No, Score = 0.455

Analysis using peptide sequences starting with any of the next four in-frame methionines downstream the first one did not result in a chloroplast targeting prediction.

26)Thaps3_4026

No introns.

>Thaps3 chr_3:2380128-2379629
TCAACAAAATCATGTCCAAGAACGCAACTACGCCCTCTTCACCTCCCGCAGCCGCAGGCA
TGGATGGAATCCTAGCCGAACGAATGGCCGAGGGAGGTGGAGCACTCTGCAACGATGAAA
ACGGATTGTGTCTAGGTTCACGTGGGGATATTGATACTAGTCAGAGTGGTTGTTATACGG
CAATTGCAAAACTTGCGAGTCAGTTGGATAACACGGTGGATGGACAGGGCACTAAGAATA
ACAACAACAACGGGGAAATGCCATTGGTGACCATTCAGACTGAAAAGGCGGCACTGTTGG
TGAAGGAATATGGGGGAAGAACGGTGGTGTTTCGTGTGCCGACAGAGGTGAATGTGAACG
GACAACAATGTGGGAGTGAACTGGTGATGAACGGGGAGCTCGACAATCTGAACGAAGGAG
TTTGATGGGGATGCATCCCGAAACAACAATCCCACTCGGTGTATATTAGGAACGGTATAT
AATACAAGAGGTTACAACAG

Translated:

MSKNATTPSSPPAAAGMDGILAERMAEGGGALCNDENGLCLGSRGDIDTSQSGCYTA
IAKLASQLDNTVDGQGTKNNNNNGEMPLVTIQTEKAALLVKEYGGRTVVFRVPTEVNVNG
QQCGSELVMNGELDNLNEGV-

Targeting:

SignalP 3.0 - NN: No

SignalP 3.0 - HMM: No, probability = 0.011

SignalP 4.1: No, D = 0.139 (D-cutoff = 0.450)

ChloroP: No, Score = 0.447

Analysis using peptide sequences starting with either of the next two in-frame methionines downstream the first one did not result in a more clear targeting prediction.

27)Thaps3_5221

Exon 1:
>Thaps3 chr_5:652777-650770
TTTCATCCTCCTTCTACACTTCATCAACGTCCGCTAGACTGCCGCCGCTGGCAACACATT
TGTCCCTCAACTCCTGACACAGCCGAAAGAGAGACTCTCGCCATGAAGCTGTTGTTTGTA
GCGTCGACGTTGATCGGCGTACTCTCGTTCACGCCCCCTCAAGTTCTGACCGACACACGT
CATCGGCCACCAGCTCTGTGTAACTCAGGCGATGCAACAAACGATGGTGGTGATGAAGTG
CACGAAGTCGATATTGCTGTCGTTGGAGCGGGGATAGGTGGTCTCTGTGCCGGCGCCATA
CTCAACACACTTTACGACAAGAAGGTTGGGGTGTATGAATCTCACTATTTAGCCGGAGGA
TGTGCACACAGTTTCAGCCGTAGTGTAAAAATTGGAGACGATGAACAGCCAACAACGTTC
ACATTTGACTCTGGGCCTACCATAGTATTGGGATGCAGCAAAGAACCGTACAATCCTCTG
CAACAAGTACTACGTGCAGTGGGGGTAGATGATCAAATAGAGTGGCTTCCTTACGACGGG
TGGGGAATGATCGAGCATCCAATGCAACCGAAGGAAAAGAGATGGAAGTTCAAAGTTGGA
CCGAATCACTTTGAGGACGGTCCTCTTCAAGTGTTTGCATCAAATCTTAATGCTCTTGAG
GAGTTCAATCAATTGAGAGAAATTACAAAGCCTCTTGTCACAGGAGCCGCTACCATTCCA
GCCATGGCCATGAGACCAGGACAATCAGCTCTAGTTCCGTTGTTGAGATATCTTCCATCA
TTGATCTCAATCATTAGCAATGGAGTTGAAGCATCGACTGGACCTTTTGCTCCCTACATG
AACGGCCCAATATTTACTGTAAAAGATCCGTGGCTACGAAGCTGGTTGAATGCGTTGGCG
TTCAGTTTGAGTGGTCTCCCCGCAGATCGTACCAGTGCTGGTGCAATGGCGTATGTGCTA
TTTGATATGCACAGAGAAGGGGCAGCGCTAGATTATCCCCGGGGAGGACTTGGAGAAGTA
GTCAAAGCATTGGTCAACGGCGTGGAGCAAAAGAGTATTGGATCAAAAGTACATCTTAGC
AGACACGTAGAAAGTATTGATACCAACGAAGAAGGAGATAGAGTCATTGGATTGACTGTT
CGTAAGAATGGAGGGAAGAAGGTCATCGTCAAGGCCAAAGAAGGTGTCGTGTGTAACGTG
CCGATGTGGTCACTCCGAAAGTTGATCAAGAATAGGAATGCACTGAGTGTTCTGGGTGGA
GACAAGGCAACTTCTTCATCAAGTGGCTTGAAAGCAAAACAATCTTGGATGACGTCTTTT

GATACAGACCCAAGTACTGGAAGAGGAAGCGTACTTCGTCCAAAACCAGCTGAGGACACG
ACAATAGAAAAAAGTCTCTTAGAGAAGTGTGACTCTGCAGAAATGACTGGCTCATTTCTT
CACCTGCATCTCGCTCTCAATGCTACTGGACTTGATCTTCAGTCTCTTGAGCCTCACTAC
ACTGTCATGGATCGTGGTTTGGAAGGCGATGGGAAAGTTATTGATGGGGTTAAGGATGAT
TCAAGCGGCGAGCTGAATATGATTGCTGTATCTAATCCTTGTGTGTTGGACAATACTTTG
GCACCAGAGGGATTTATCATCATGCATGCCTATGGTGCAGGTAACGAGCCTTTCGAGATA
TGGAAACCACCAACTGCAAGTAAAGGCAATGCTTCACCAAATACTGCAGGAGAAGGAGAA
ATTATTGGAGGGGAACGATGCTCACCATCGACGTACCAGGCATTGAAAGATAGCCGATCG
AAGGTACTATGGAGAGCTGTGGAGTCTGTTATACCTGACGCACGTGAACGTACTGTGCTT
GCTCTCATCGGATCTCCTCGAACACACGAACGATTTCTACGTCGTCCATGCGGCTCGTAC
GGTGCAGCGTTTGAGGATTGTTTGAAGGACGGAAGCACTCCAATATCTAACTTGGTCTTA
TCTGGTGACGGTGTCTTTCCTGGTATTG

Exon 2:
>Thaps3 chr_5:650680-650446
GCATTCCTGCTGTAGCACTCAACGGGGCTAGCGCAGCGAATGGATTCGTTGGCATATTTG
ATCAGTGGAGATGTATGGATTATCTTAAGGCCAAAGGAATCATTGCCTAGATAGCCGAGA
AGTTGGACTTCGTCGGCAGCAGTATTTTAGGGTTCCGTGTGCCTTCTTCAACTGAGTATC
GCCCATTGTACTCAGTCGCTTTCAAGCTACTGTCTCTTTCAACGAGTTCACAAAT

Translated:

MKLLFVASTLIGVLSFTPPQVLTDTR
HRPPALCNSGDATNDGGDEVHEVDIAVVGAGIGGLCAGAILNTLYDKKVGVYESHYLAGG
CAHSFSRSVKIGDDEQPTTFTFDSGPTIVLGCSKEPYNPLQQVLRAVGVDDQIEWLPYDG
WGMIEHPMQPKEKRWKFKVGPNHFEDGPLQVFASNLNALEEFNQLREITKPLVTGAATIP
AMAMRPGQSALVPLLRYLPSLISIISNGVEASTGPFAPYMNGPIFTVKDPWLRSWLNALA
FSLSGLPADRTSAGAMAYVLFDMHREGAALDYPRGGLGEVVKALVNGVEQKSIGSKVHLS
RHVESIDTNEEGDRVIGLTVRKNGGKKVIVKAKEGVVCNVPMWSLRKLIKNRNALSVLGG
DKATSSSSGLKAKQSWMTSFDTDPSTGRGSVLRPKPAEDTTIEKSLLEKCDSAEMTGSFL
HLHLALNATGLDLQSLEPHYTVMDRGLEGDGKVIDGVKDDSSGELNMIAVSNPCVLDNTL
APEGFIIMHAYGAGNEPFEIWKPPTASKGNASPNTAGEGEIIGGERCSPSTYQALKDSRS
KVLWRAVESVIPDARERTVLALIGSPRTHERFLRRPCGSYGAAFEDCLKDGSTPISNLVL
SGDGVFPGIGIPAVALNGASAANGFVGIFDQWRCMDYLKAKGIIA-

Targeting:

SignalP 3.0 - NN: Yes, Cleavage Site VLS-FT

SignalP 3.0 - HMM: Yes, Cleavage Site VLT-DT, Signal Peptide Probability = 0.999

SignalP 4.1: Yes, Cleavage Site VLS-FT, D = 0.635 (D-cutoff = 0.500)

ChloroP: No, Score = 0.460

Analysis using peptide sequences starting with either of the next two in-frame methionines downstream the first one did not result in a more clear different targeting prediction.

APPENDIX 2.D ADDITIONAL SEQUENCE-BASED ANALYSES

1)PDS

Peptide Alignment (With the JGI-Predicted Intron for Thaps3_bd_1474)

```
PDS1         MIITNFILSTVLATSMAFQPHTPILSKPSFSNRVHRSPKIGSSNLVMKDFPKPNVEDTDN    60
bd_1474      ------------------------------------------------------------    0


PDS1         YRYAEAMSTSFKTSLRVTNDSQKKKVAIIGGGLSGLSCAKYLSDAGHEPTVYEARDVLGG    120
bd_1474      ------------------------------------------------------------    0


PDS1         KVSAWQDEDGDWIETGLHIFFGAYPNVMNMFAELGIHDRLQWKIHQMIFAMQELPGEFTT    180
bd_1474      ------------------------------------------------------------    0


PDS1         FDFIPGIPAPFNFGLAILMNQKMLTLGEKIQTAPPLLPMLIEGQSFIDAQDELSVTQFMR    240
bd_1474      ------------------MNQKMLTLGEKIQTAPPLLPMLIEGQSFIDAQDELSVTQFMR    42
                              ****************************************


PDS1         KYGMPERINEEVFIAMAKALDFIDPDKLSMTVVLTAMNRFLNESNGLQMAFLDGNQPDRW    300
bd_1474      KYGMPERINEEVFIAMAKALDFIDPDKLSMTVVLTAMNRFLNESNGLQMAFLDGNQPDRW    102
             ************************************************************


PDS1         CTPTKEYVEARGGKVKLNSPIKEIVTNDDGTINHLLLRSGEKIVADEYVSAMPVDIVKRM    360
bd_1474      CTPTKEYVEARGGKVKLNSPIKEIVTNDDGTINHLLLRSGEKIVADEYVSAMPVDIVKRM    162
             ************************************************************


PDS1         LPTTWQTMPYFRQLDELEGIPVINLHMWFDRKLKAVDHLCFSRSPLLSVYADMSVTCKEY    420
bd_1474      LPTTWQTMPYFRQLDELEGIPVINLHMWFDRKLKAVDHLCFSRSPLLSVYADMSVTCKEY    222
             ************************************************************


PDS1         EDPNKSMLELVFAPCSPIAGGNVNWIGKSDEEIIDATMGELARLFPTEIANDDKWPATKM    480
bd_1474      EDPNKSMLELVFAPCSPIAGGNVNWIGKSDEEIIDATMGELARLFPTEIANDDKWPATKM    282
             ************************************************************


PDS1         QGPNGQAKLEKYAVVKVPRSVYAAIPGE--------------------------------    508
bd_1474      QGPNGQAKLEKYAVVKVPRSVYAAIPGRNKYRPSQTSPIPHFTMAGCYTSQKFLGSMEGA    342
             ***************************.

PDS1         ------------------------------------------------------------    508
bd_1474      TLAGKLAAEVIANRALGNADKPVKEIQQHIIDSASKHVVKEPVGVKGEGAIAFGGGYTVG    402


PDS1         ----------------------          508
bd_1474      KKEEDLLRESDPAQYELAVAK           423
```

Peptide Alignment (Disregarding the JGI-Predicted Intron for Thaps3_bd_1474)

```
bd_1474_No_Intron      ------------------------------------------------------------   0
PDS1                   MIITNFILSTVLATSMAFQPHTPILSKPSFSNRVHRSPKIGSSNLVMKDFPKPNVEDTDN   60


bd_1474_No_Intron      ------------------------------------------------------------   0
PDS1                   YRYAEAMSTSFKTSLRVTNDSQKKKVAIIGGGLSGLSCAKYLSDAGHEPTVYEARDVLGG   120


bd_1474_No_Intron      ------------------------------------------------------------   0
```

```
PDS1                KVSAWQDEDGDWIETGLHIFFGAYPNVMNMFAELGIHDRLQWKIHQMIFAMQELPGEFTT   180


bd_1474_No_Intron   ------------------MNQKMLTLGEKIQTAPPLLPMLIEGQSFIDAQDELSVTQFMR   42
PDS1                FDFIPGIPAPFNFGLAILMNQKMLTLGEKIQTAPPLLPMLIEGQSFIDAQDELSVTQFMR   240
                                      ****************************************


bd_1474_No_Intron   KYGMPERINEEVFIAMAKALDFIDPDKLSMTVVLTAMNRFLNESNGLQMAFLDGNQPDRW   102
PDS1                KYGMPERINEEVFIAMAKALDFIDPDKLSMTVVLTAMNRFLNESNGLQMAFLDGNQPDRW   300
                    ************************************************************


bd_1474_No_Intron   CTPTKEYVEARGGKVKLNSPIKEIVTNDDGTINHLLLRSGEKIVADEYVSAMPVDIVKRM   162
PDS1                CTPTKEYVEARGGKVKLNSPIKEIVTNDDGTINHLLLRSGEKIVADEYVSAMPVDIVKRM   360
                    ************************************************************


bd_1474_No_Intron   LPTTWQTMPYFRQLDELEGIPVINLHMWFDRKLKAVDHLCFSRSPLLSVYADMSVTCKEY   222
PDS1                LPTTWQTMPYFRQLDELEGIPVINLHMWFDRKLKAVDHLCFSRSPLLSVYADMSVTCKEY   420
                    ************************************************************


bd_1474_No_Intron   EDPNKSMLELVFAPCSPIAGGNVNWIGKSDEEIIDATMGELARLFPTEIANDDKWPATKM   282
PDS1                EDPNKSMLELVFAPCSPIAGGNVNWIGKSDEEIIDATMGELARLFPTEIANDDKWPATKM   480
                    ************************************************************


bd_1474_No_Intron   QGPNGQAKLEKYAVVKVPRSVYAAIPGE      310
PDS1                QGPNGQAKLEKYAVVKVPRSVYAAIPGE      508
                    ****************************
```

Genomic Alignment (Extended Sequences Contain 10kb downstream the JGI-Predicted Gene Models)

The abrupt start of sequence differences is highlighted in red.

```
PDS1_Extended       GCAAAGCGTTCCTCTTCGGCGTCAGCCTCCTCCTTTCTTTCACTGTCAGTCAGTTGTGTT    60
PDS1                GCAAAGCGTTCCTCTTCGGCGTCAGCCTCCTCCTTTCTTTCACTGTCAGTCAGTTGTGTT    60
bd_1474_Extended    ------------------------------------------------------------    0
bd_1474             ------------------------------------------------------------    0


PDS1_Extended       GAACACTTCTGCTTCTTCATTCATCTCCTCATTACTACCACTGCTATTGCAACTGGCTGC    120
PDS1                GAACACTTCTGCTTCTTCATTCATCTCCTCATTACTACCACTGCTATTGCAACTGGCTGC    120
bd_1474_Extended    ------------------------------------------------------------    0
bd_1474             ------------------------------------------------------------    0


PDS1_Extended       CGTATCTGACTCTCCTTCAATTCGTCACTCCCATAGCGCCACGTTCACCATCATTTATCA    180
PDS1                CGTATCTGACTCTCCTTCAATTCGTCACTCCCATAGCGCCACGTTCACCATCATTTATCA    180
bd_1474_Extended    ------------------------------------------------------------    0
bd_1474             ------------------------------------------------------------    0


PDS1_Extended       CCATGATCATTACAAATTTCATCCTCTCCACCGTCCTAGCGACATCAATGGCCTTTCAAC    240
PDS1                CCATGATCATTACAAATTTCATCCTCTCCACCGTCCTAGCGACATCAATGGCCTTTCAAC    240
bd_1474_Extended    ------------------------------------------------------------    0
bd_1474             ------------------------------------------------------------    0


PDS1_Extended       CACACACACCCATCCTCTCCAAACCATCCTTCTCCAACCGTGTCCATCGCTCCCCCAAAA    300
PDS1                CACACACACCCATCCTCTCCAAACCATCCTTCTCCAACCGTGTCCATCGCTCCCCCAAAA    300
bd_1474_Extended    ------------------------------------------------------------    0
bd_1474             ------------------------------------------------------------    0


PDS1_Extended       TCGGCTCTTCCAACCTCGTTATGAAGGACTTTCCGAAACCAAATGTCGAAGATACAGACA    360
PDS1                TCGGCTCTTCCAACCTCGTTATGAAGGACTTTCCGAAACCAAATGTCGAAGATACAGACA    360
bd_1474_Extended    ------------------------------------------------------------    0
bd_1474             ------------------------------------------------------------    0
```

```
PDS1_Extended        ACTATCGCTACGCAGAGGCCATGTCCACTAGCTTCAAGACGTCTCTCCGAGTGACGAATG    420
PDS1                 ACTATCGCTACGCAGAGGCCATGTCCACTAGCTTCAAGACGTCTCTCCGAGTGACGAATG    420
bd_1474_Extended     ------------------------------------------------------------    0
bd_1474              ------------------------------------------------------------    0

PDS1_Extended        ATTCACAGAAGAAGAAGGTGGCTATCATTGGAGGAGGATTATCAGGTCTGTCTTGTGCCA    480
PDS1                 ATTCACAGAAGAAGAAGGTGGCTATCATTGGAGGAGGATTATCAGGTCTGTCTTGTGCCA    480
bd_1474_Extended     ------------------------------------------------------------    0
bd_1474              ------------------------------------------------------------    0

PDS1_Extended        AGTACCTCTCCGATGCCGGGCATGAACCCACCGTATACGAAGCACGTGATGTACTCGGAG    540
PDS1                 AGTACCTCTCCGATGCCGGGCATGAACCCACCGTATACGAAGCACGTGATGTACTCGGAG    540
bd_1474_Extended     ------------------------------------------------------------    0
bd_1474              ------------------------------------------------------------    0

PDS1_Extended        GAAAGGTGTCAGCGTGGCAAGATGAAGATGGAGACTGGATCGAAACAGGTCTTCACATCT    600
PDS1                 GAAAGGTGTCAGCGTGGCAAGATGAAGATGGAGACTGGATCGAAACAGGTCTTCACATCT    600
bd_1474_Extended     ------------------------------------------------------------    0
bd_1474              ------------------------------------------------------------    0

PDS1_Extended        TCTTCGGAGCATACCCCAACGTTATGAACATGTTCGCTGAGCTTGGCATCCACGATAGGC    660
PDS1                 TCTTCGGAGCATACCCCAACGTTATGAACATGTTCGCTGAGCTTGGCATCCACGATAGGC    660
bd_1474_Extended     ------------------------------------------------------------    0
bd_1474              ------------------------------------------------------------    0

PDS1_Extended        TTCAGTGGAAGATTCACCAAATGATTTTCGCAATGCAGGAACTTCCCGGAGAGTTCACTA    720
PDS1                 TTCAGTGGAAGATTCACCAAATGATTTTCGCAATGCAGGAACTTCCCGGAGAGTTCACTA    720
bd_1474_Extended     ------------------------------------------------------------    0
bd_1474              ------------------------------------------------------------    0

PDS1_Extended        CCTTTGATTTCATCCCTGGTATTCCAGCTCCGTTCAACTTTGGATTGGCCATTCTTATGA    780
PDS1                 CCTTTGATTTCATCCCTGGTATTCCAGCTCCGTTCAACTTTGGATTGGCCATTCTTATGA    780
bd_1474_Extended     -----------------------------------------------------CTTATGA    7
bd_1474              -----------------------------------------------------CTTATGA    7
                                                                          *******

PDS1_Extended        ATCAAAAGATGTTGACGTTGGGTGAAAAAATTCAGACCGCTCCTCCTCTTCTTCCTATGC    840
PDS1                 ATCAAAAGATGTTGACGTTGGGTGAAAAAATTCAGACCGCTCCTCCTCTTCTTCCTATGC    840
bd_1474_Extended     ATCAAAAGATGTTGACGTTGGGTGAAAAAATTCAGACCGCTCCTCCTCTTCTTCCTATGC    67
bd_1474              ATCAAAAGATGTTGACGTTGGGTGAAAAAATTCAGACCGCTCCTCCTCTTCTTCCTATGC    67
                     ************************************************************

PDS1_Extended        TTATTGAGGGACAGTCATTCATTGATGCTCAGGATGAGTTGAGTGTGACGCAGTTCATGA    900
PDS1                 TTATTGAGGGACAGTCATTCATTGATGCTCAGGATGAGTTGAGTGTGACGCAGTTCATGA    900
bd_1474_Extended     TTATTGAGGGACAGTCATTCATTGATGCTCAGGATGAGTTGAGTGTGACGCAGTTCATGA    127
bd_1474              TTATTGAGGGACAGTCATTCATTGATGCTCAGGATGAGTTGAGTGTGACGCAGTTCATGA    127
                     ************************************************************

PDS1_Extended        GGAAGTACGGTATGCCTGAGAGAATCAACGAGGAGGTGTTTATTGCGATGGCCAAGGCGT    960
PDS1                 GGAAGTACGGTATGCCTGAGAGAATCAACGAGGAGGTGTTTATTGCGATGGCCAAGGCGT    960
bd_1474_Extended     GGAAGTACGGTATGCCTGAGAGAATCAACGAGGAGGTGTTTATTGCGATGGCCAAGGCGT    187
bd_1474              GGAAGTACGGTATGCCTGAGAGAATCAACGAGGAGGTGTTTATTGCGATGGCCAAGGCGT    187
                     ************************************************************

PDS1_Extended        TGGACTTTATTGATCCTGATAAGTTGAGTATGACTGTGGTGCTTACGGCTATGAACAGGT    1020
PDS1                 TGGACTTTATTGATCCTGATAAGTTGAGTATGACTGTGGTGCTTACGGCTATGAACAGGT    1020
bd_1474_Extended     TGGACTTTATTGATCCTGATAAGTTGAGTATGACTGTGGTGCTTACGGCTATGAACAGGT    247
bd_1474              TGGACTTTATTGATCCTGATAAGTTGAGTATGACTGTGGTGCTTACGGCTATGAACAGGT    247
                     ************************************************************

PDS1_Extended        TCTTGAATGAGAGTAATGGACTTCAGATGGCATTCTTGGATGGAAATCAGCCTGATAGGT    1080
PDS1                 TCTTGAATGAGAGTAATGGACTTCAGATGGCATTCTTGGATGGAAATCAGCCTGATAGGT    1080
bd_1474_Extended     TCTTGAATGAGAGTAATGGACTTCAGATGGCATTCTTGGATGGAAATCAGCCTGATAGGT    307
```

```
bd_1474              TCTTGAATGAGAGTAATGGACTTCAGATGGCATTCTTGGATGGAAATCAGCCTGATAGGT    307
                     ************************************************************

PDS1_Extended        GGTGCACTCCCACGAAGGAGTATGTGGAAGCACGCGGAGGAAAGGTCAAATTGAACTCTC    1140
PDS1                 GGTGCACTCCCACGAAGGAGTATGTGGAAGCACGCGGAGGAAAGGTCAAATTGAACTCTC    1140
bd_1474_Extended     GGTGCACTCCCACCAAGGAGTATGTGGAAGCACGCGGAGGAAAGGTCAAATTGAACTCTC    367
bd_1474              GGTGCACTCCCACCAAGGAGTATGTGGAAGCACGCGGAGGAAAGGTCAAATTGAACTCTC    367
                     ************ ***********************************************

PDS1_Extended        CCATTAAGGAGATTGTGACCAACGACGATGGAACTATCAATCACCTTCTCCTTCGATCTG    1200
PDS1                 CCATTAAGGAGATTGTGACCAACGACGATGGAACTATCAATCACCTTCTCCTTCGATCTG    1200
bd_1474_Extended     CCATTAAGGAGATTGTGACCAACGACGATGGAACTATCAATCACCTTCTCCTTCGATCTG    427
bd_1474              CCATTAAGGAGATTGTGACCAACGACGATGGAACTATCAATCACCTTCTCCTTCGATCTG    427
                     ************************************************************

PDS1_Extended        GCGAGAAGATTGTGGCCGATGAATACGTCTCTGCCATGCCCGTGGACATCGTCAAACGTA    1260
PDS1                 GCGAGAAGATTGTGGCCGATGAATACGTCTCTGCCATGCCCGTGGACATCGTCAAACGTA    1260
bd_1474_Extended     GCGAGAAGATTGTGGCCGATGAATACGTCTCTGCCATGCCCGTGGACATCGTCAAACGTA    487
bd_1474              GCGAGAAGATTGTGGCCGATGAATACGTCTCTGCCATGCCCGTGGACATCGTCAAACGTA    487
                     ************************************************************

PDS1_Extended        TGCTTCCCACAACGTGGCAGACTATGCCCTACTTCCGTCAGCTTGACGAACTTGAGGGCA    1320
PDS1                 TGCTTCCCACAACGTGGCAGACTATGCCCTACTTCCGTCAGCTTGACGAACTTGAGGGCA    1320
bd_1474_Extended     TGCTTCCCACAACGTGGCAGACTATGCCCTACTTCCGTCAGCTTGACGAACTTGAGGGCA    547
bd_1474              TGCTTCCCACAACGTGGCAGACTATGCCCTACTTCCGTCAGCTTGACGAACTTGAGGGCA    547
                     ************************************************************

PDS1_Extended        TCCCTGTTATCAACTTGCACATGTGGTTCGATCGTAAGTTGAAAGCAGTCGACCATCTTT    1380
PDS1                 TCCCTGTTATCAACTTGCACATGTGGTTCGATCGTAAGTTGAAAGCAGTCGACCATCTTT    1380
bd_1474_Extended     TCCCTGTTATCAACTTGCACATGTGGTTCGATCGTAAGTTGAAAGCAGTCGACCATCTTT    607
bd_1474              TCCCTGTTATCAACTTGCACATGTGGTTCGATCGTAAGTTGAAAGCAGTCGACCATCTTT    607
                     ************************************************************

PDS1_Extended        GCTTCAGTCGCTCCCCACTCCTTTCCGTCTACGCCGACATGTCCGTCACATGCAAGGAGT    1440
PDS1                 GCTTCAGTCGCTCCCCACTCCTTTCCGTCTACGCCGACATGTCCGTCACATGCAAGGAGT    1440
bd_1474_Extended     GCTTCAGTCGCTCCCCACTCCTTTCCGTCTACGCTGACATGTCCGTCACATGCAAGGAGT    667
bd_1474              GCTTCAGTCGCTCCCCACTCCTTTCCGTCTACGCTGACATGTCCGTCACATGCAAGGAGT    667
                     ********************************** ************************

PDS1_Extended        ACGAAGATCCCAACAAGTCCATGTTGGAATTGGTCTTTGCTCCCTGCTCTCCTATTGCCG    1500
PDS1                 ACGAAGATCCCAACAAGTCCATGTTGGAATTGGTCTTTGCTCCCTGCTCTCCTATTGCCG    1500
bd_1474_Extended     ACGAAGATCCCAACAAGTCCATGTTGGAATTGGTCTTTGCTCCCTGCTCTCCTATTGCCG    727
bd_1474              ACGAAGATCCCAACAAGTCCATGTTGGAATTGGTCTTTGCTCCCTGCTCTCCTATTGCCG    727
                     ************************************************************

PDS1_Extended        GAGGAAATGTCAACTGGATTGGAAAGTCAGATGAGGAAATCATTGATGCTACCATGGGTG    1560
PDS1                 GAGGAAATGTCAACTGGATTGGAAAGTCAGATGAGGAAATCATTGATGCTACCATGGGTG    1560
bd_1474_Extended     GAGGAAATGTCAACTGGATTGGAAAGTCAGATGAGGAAATCATTGATGCTACCATGGGTG    787
bd_1474              GAGGAAATGTCAACTGGATTGGAAAGTCAGATGAGGAAATCATTGATGCTACCATGGGTG    787
                     ************************************************************

PDS1_Extended        AGCTTGCTCGCCTTTTCCCTACCGAGATTGCGAATGATGATAAGTGGCCTGCTACGAAGA    1620
PDS1                 AGCTTGCTCGCCTTTTCCCTACCGAGATTGCGAATGATGATAAGTGGCCTGCTACGAAGA    1620
bd_1474_Extended     AGCTTGCTCGCCTTTTCCCTACCGAGATTGCGAATGATGATAAGTGGCCTGCTACGAAGA    847
bd_1474              AGCTTGCTCGCCTTTTCCCTACCGAGATTGCGAATGATGATAAGTGGCCTGCTACGAAGA    847
                     ************************************************************

PDS1_Extended        TGCAGGGACCTAATGGACAGGCAAAGCTTGAGAAGTATGCTGTTGTGAAGGTGCCAAGGA    1680
PDS1                 TGCAGGGACCTAATGGACAGGCAAAGCTTGAGAAGTATGCTGTTGTGAAGGTGCCAAGGA    1680
bd_1474_Extended     TGCAGGGACCTAATGGACAGGCAAAGCTTGAGAAGTATGCTGTTGTGAAGGTGCCAAGGA    907
bd_1474              TGCAGGGACCTAATGGACAGGCAAAGCTTGAGAAGTATGCTGTTGTGAAGGTGCCAAGGA    907
                     ************************************************************

PDS1_Extended        GTGTGTATGCTGCCATTCCTGGTGAGTGAAAAATAGTGTGCGGTGAATCTTCGTCGTCTA    1740
PDS1                 GTGTGTATGCTGCCATTCCTGGTGAGTGAAAAATAGTGTGCGGTGAATCTTCGTCGTCTA    1740
bd_1474_Extended     GTGTGTATGCTGCCATTCCTGGTGAGTGAAAAAGAGTGTGCGGTGAATCTTCGTCGTCTA    967
bd_1474              GTGTGTATGCTGCCATTCCTGGTGAGTGAAAAAGAGTGTGCGGTGAATCTTCGTCGTCTA    967
                     ********************************** ************************

PDS1_Extended        CCTTGGCCAACTACTGACTCTTGTTCTCTTTAAATCATATCAGGACGTAACAAATACCGC    1800
PDS1                 CCTTGGCCAACTACTGACTCTTGTTCTCTTTAAATCATATCAGGACGTAACAAATACCGC    1800
```

```
bd_1474_Extended      CCTTGGCCAACTACTGACTCTTGTTCTCTTTAAATCATATCAGGACGTAACAAATACCGC   1027
bd_1474               CCTTGGCCAACTACTGACTCTTGTTCTCTTTAAATCATATCAGGACGTAACAAATACCGC   1027
                      ************************************************************

PDS1_Extended         CCCAGTCAGACCTCCCCCATCCCACACTTCACCATGGCTGGATGCTATACCTCACAAAAG   1860
PDS1                  CCCAGTCAGACCTCCCCCATCCCACACTTCACCATGGCTGGATGCTATACCTCACAAAAG   1860
bd_1474_Extended      CCCAGTCAGACCTCCCCCATCCCACACTTCACCATGGCTGGATGCTATACCTCACAAAAG   1087
bd_1474               CCCAGTCAGACCTCCCCCATCCCACACTTCACCATGGCTGGATGCTATACCTCACAAAAG   1087
                      ************************************************************

PDS1_Extended         TTCCTCGGATCCATGGAGGGTGCCACCCTCGCCGGGGAAGCTTGCTGCCGAGGTCATTGCC   1920
PDS1                  TTCCTCGGATCCATGGAGGGTGCCACCCTCGCCGGGGAAGCTTGCTGCCGAGGTCATTGCC   1920
bd_1474_Extended      TTCCTCGGATCCATGGAGGGTGCCACCCTCGCCGGGGAAGCTTGCTGCCGAGGTCATTGCC   1147
bd_1474               TTCCTCGGATCCATGGAGGGTGCCACCCTCGCCGGGGAAGCTTGCTGCCGAGGTCATTGCC   1147
                      ************************************************************

PDS1_Extended         AACCGTGCCCTCGGAAATGCGGATAAGCCAGTCAAGGAGATTCAGCAACACATTATCGAC   1980
PDS1                  AACCGTGCCCTCGGAAATGCGGATAAGCCAGTCAAGGAGATTCAGCAACACATTATCGAC   1980
bd_1474_Extended      AACCGTGCCCTCGGAAATGCGGATAAGCCAGTCAAGGAGATTCAGCAACACATTATCGAC   1207
bd_1474               AACCGTGCCCTCGGAAATGCGGATAAGCCAGTCAAGGAGATTCAGCAACACATTATCGAC   1207
                      ************************************************************

PDS1_Extended         TCGGCTAGTAAGCATGTTGTGAAGGAGCCAGTGGGTGTGAAGGGAGAGGGAGCGATTGCA   2040
PDS1                  TCGGCTAGTAAGCATGTTGTGAAGGAGCCAGTGGGTGTGAAGGGAGAGGGAGCGATTGCA   2040
bd_1474_Extended      TCGGCTAGTAAGCATGTTGTGAAGGAGCCAGTGGGTGTGAAGGGAGAGGGAGCGATTGCA   1267
bd_1474               TCGGCTAGTAAGCATGTTGTGAAGGAGCCAGTGGGTGTGAAGGGAGAGGGAGCGATTGCA   1267
                      ************************************************************

PDS1_Extended         TTTGGAGGGGGGTATACTGTTGGAAAGAAGGAGGAGGATTTGTTGAGGGAGTCGGATCCT   2100
PDS1                  TTTGGAGGGGGGTATACTGTTGGAAAGAAGGAGGAGGATTTGTTGAGGGAGTCGGATCCT   2100
bd_1474_Extended      TTTGGAGGGGGGTATACTGTTGGAAAGAAGGAGGAGGATTTGTTGAGGGAGTCGGATCCT   1327
bd_1474               TTTGGAGGGGGGTATACTGTTGGAAAGAAGGAGGAGGATTTGTTGAGGGAGTCGGATCCT   1327
                      ************************************************************

PDS1_Extended         GCTCAGTATGAGTTGGCAGTAGCCAAGTAAGGAAGAGTATTATTAAAGTAAACAGCTAGA   2160
PDS1                  GCTCAGTATGAGTTGGCAGTAGCCAAGTAAGGAAGAGTATTATTAAAGTAAACAGCTAGA   2160
bd_1474_Extended      GCTCAGTATGAGTTGGCAGTAGCCAAGTAAGGAAGAGTATTATTAAAGTAAACAGCTAGA   1387
bd_1474               GCTCAGTATGAGTTGGCAGTAGCCAAGTAA------------------------------   1357
                      ******************************

PDS1_Extended         TTGATTTTGAAGGAGTTTTGTATGTAGCTTAGTACCTTTAAGGTACTTGATGTACTTACT   2220
PDS1                  TTGATTTTGAA------------------------------------------------   2171
bd_1474_Extended      TTGATTTTGAAGGAGTTTTGTATGTAACTTAGTACCTTTAAGGTACTTGATGTTCTTACT   1447
bd_1474               ------------------------------------------------------------   1357

PDS1_Extended         TTCCGCTTGGAGAGCCATTTCTTTGATTATATGACAGAATGGAGTTTTCCCTTCATCCTT   2280
PDS1                  ------------------------------------------------------------   2171
bd_1474_Extended      TTCCGCTTGGAGAGCCATTTCTTTGATTTTACGACAGAATTGAGTTTTCCCTTCATCCTT   1507
bd_1474               ------------------------------------------------------------   1357

PDS1_Extended         CTCGTTGATTGATTTTGGATCAGATGATATTTATGCATCTCTCACCTGGTGCCACTAACG   2340
PDS1                  ------------------------------------------------------------   2171
bd_1474_Extended      CTCGTTGATTGATTTTGGATCAGATGATATTTATGCATCTCTCACCTGGTGCCACTAACG   1567
bd_1474               ------------------------------------------------------------   1357

PDS1_Extended         ATACATACGAAAGAACCTCCCCCCACCTTGCCACACCAACTCATTCATCTCATCTCCTTA   2400
PDS1                  ------------------------------------------------------------   2171
bd_1474_Extended      ATACATACGAAAGAACCTCCC-CCACCTTGCCACACCAACTCATTCATCTCATCTCCTTA   1626
bd_1474               ------------------------------------------------------------   1357

PDS1_Extended         TTTGATTTTGTGCTCCGTAAACAGAACTCTCCTCCTGACAATCGGATCAAATTTTACGAA   2460
PDS1                  ------------------------------------------------------------   2171
bd_1474_Extended      TTTGATTTTGTGCTCCGTAAACAGAACTCTCCTCCTGACAATTGGATCAAATTTTACGAA   1686
bd_1474               ------------------------------------------------------------   1357

PDS1_Extended         CTTGAACTTCTCTGGTGTCTTCGATACGTTTCGTCGTGTGGTGTAGAAGAAGCCAGTCCC   2520
```

```
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    CTTGAACTTCTCTGGTGTCTTCGATACGTTTCGTCGTGTGGTGTAGAAGAAGCCAGTCCC    1746
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       CGCCGAGCTTAAGAGCTGCGTTGTTAACGACGTGAATGGAAAGAGGAGTCAGTCAGTGTT    2580
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    CGCCGAGCTTAAGAGCTGCGTTGTTAACGACGTGAATGGAAAGAGGAGTCAGTCAGTGTT    1806
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       GGATGAGATGATATGCAGCATAGCCACACCTCTTCAATCATACAAAGGCTGATTTGCACA    2640
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    GGATGAGATGATATGCAGCATAGCCACAACTCTTCAATCATACAAAGGCTGATTTGCACA    1866
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       CAACGTGTACCGTCACCTCATCTGACGGAACAGCACCTCCCACTCACTTTGATGGGGATC    2700
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    CAACGTGTACCGTCACCTCATCTGACGGAACAGCACCTCCCACTCACTTTGATGGGGATC    1926
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       GTCTTGCCCTTCCCTCTGGCCATGTGAAGAATAGCGGTCGGCTGCATGGTATTTTACTCT    2760
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    GTCTTGCCCTTCCCTCTGGCCATGTGAAGAATAGCGGTCGGCTGCATGGTATTTTACTCT    1986
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       TCTCGATAATGCAAACGTTTTCTAAGATGGGTTGTTGCCGGTTGTGGTGAAGGGGTGACG    2820
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    TCTCGATACTGCAGACGTTTTCTAAGATGGGTTGTTGCCGGTTGTGGTGAAGGGGTGGCG    2046
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       GTGGTGGTGGTTGTGGTTGTTGTTTTGTGTGCGTGACGAGTTGTGATGTTGTGGTTTTGT    2880
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    GTGGTGGTCGTTGTGGTTGTTGTTTTGTGTGCGTGACGAGTTGTGATGTTGTGGTTTTGT    2106
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       GAACAGTTCAGTAACTTGTTGTCGTGACGGACTAACGTGATTCCGTGGATACCAAGAAGC    2940
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    GAACAGTTCTGTAACGTGTTGTCGTGACGGACTAACGTGATTCCGTGGATACCAAGAAGC    2166
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       ACTAACATGGTAACCCCC--CCCCCAGACCAACAAATCATTTCATCCCCACCAACTCCAT    2998
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    ACTAACATAGTAACCCCACCCCCCCAGACCAACAAATCATTTCATCCCCACCAACTCCAT    2226
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       CACACTCCCTCATCGATACATCCAGCACAAACCATGATAGTTGATAGTCTCCATTGATTC    3058
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    CACACTCCCTCATCGATACATCCAGCACAAACCATGATAGTTGATAGTCTCCATTGATTC    2286
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       CTCGTCGCAGTCTGCTACAGCGATGCCTCCACAAACAGCCGTCACGATGATAAACATTGT    3118
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    CTCGTCGCAGTCTGCTACAGCGATGCCTCCACAAACAGCCGTCACATTGATAAACATTGT    2346
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       ACTCCCACTCCAGACTGGGTGGAGTCGACAGTGAAGTCAAGACGATGATGAATGTTGGCT    3178
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    ACTCCCACTCCAGACTGGGTGGAGTCGACAGTGAAGTCAAGACGATGATGAATGTTGGCT    2406
bd_1474             ------------------------------------------------------------    1357
```

```
PDS1_Extended        GCAAACGCACAACCCCGAGCCCAAGTGAACGATGGCACTACAATCATCCCGCCGCCATGC    3238
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GCAAACGCACAACCCCGAGCCCAAGTGAACGATGGCACTACAATCATCCCGCCGCCATGC    2466
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        CTTTGAAACCAATTCACTAGACCACTGCGAAACACCACTTTGTGCCTACGAAAATGTTCA    3298
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     CTTTGAAACCAATTCACTAGACCACTGCGAAACACCACTTTGTGCCTACGAAAATGTTCA    2526
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        ACCAGTCCTCGAGATGATGGCGAAGCATCTTCACGTCCAACCATCAATGTTACGCATCTG    3358
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     AACAGTCCTCGAGATGATGGCGAAGCATCTTCACGTCCAACCATCAAAGTTACGCATCTG    2586
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        GGATCCGTATTATTGCGATGGGACTGTCAAGCAACATCTAGCCTCTTTGGGATACGATCG    3418
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GGATCCGTATTATTGCGATGGGACTGTCAAGCAACATCTAGCCTCTTTGGGATACGATCG    2646
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        TGTGATTAACGAAAACTTTGACTTTTACAAGCGAGTAGAGGACAACACTATCCCCGAGCA    3478
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     TGTGATTAACGAAAACGTTGACTTTTACAAGCGAGTAGAGGACAACACTATCCCCGAGCA    2706
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        CGATGTCTTGTTGACGAATCCTCCGTACAGCGGCGATCACATCGAACGATTGCTTAAGTT    3538
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     CGATGTCTTGTTGACGAATCCTCCGTACAGCGGCGATCACATCGAACGATTGCTTAAGTT    2766
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        TGTTACGACGGTGAATGATAAGCCATTTTGTCTACTAATGCCAAATTGGGTTGCGAGAAA    3598
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     TGTTACGACGGTGAATGATAAGCCATTTTGTCTACTAATGCCAAATTGGGTTGCGAGAAA    2826
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        GAAGGAGTACAAATCCATCATTGGTAAAACGAATCTGTTTTACGTATCCCCCATCGAGGT    3658
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GAAGGAGTACAAATCCATCATTGGTAAAACGAATCTGTTTTACGTATCCCCCATAGAGGT    2886
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        ATACACGTATGCTATGCCAACTTGGAATTCGAAACCGGAACACGTCGACGAGGAGACGGG    3718
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     ATACACGTATGCTATGCCAACTTGGAATTCGAAACCGGAACACGTCGACGAGGAGACGGG    2946
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        AAAGACGACTCCCTATTTGAGTTCTTGGTATGTATCGTTAAGAAGCAATAGTGAGGCAAC    3778
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     AAAGACGACACCCTATTTGAGTTCTTGGTATGTATCGTTAAGAAGCAATAGTGAGGCAAC    3006
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        GAGTAGGATAGAGAATAAGCTTGATTCTATTGCAAAACGGCAGCAACCACCTGTTTGGGT    3838
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GGGTAGGATAGAGAATAAGCTTGATTCTATTGCAAAACGGCAGCAACCACCTGTTTGGGT    3066
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        TGTAGCCAAGACGGTCAAAGGGCTAAAGTGGAAGATTCAAAAGGTGAAAGAAAAGAAGAG    3898
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     TGTAGCCAAGACGGTCAAATCGACAGGAGCACCCAATTCCTTGACAACATTGTTGTGTGT    3126
bd_1474              ------------------------------------------------------------    1357
```

```
PDS1_Extended        GTGATTGGAGATTTT-----CTCCTCGGAGATGCTAAATCGCTTAGCAATGGAAGTGAGA    3953
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     CTCAATGGGAGCGATGAAATCGCATGGGACAATCTTCGTGCCTTGATGCAATAATTGATA    3186
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        A-------------CTGAATTGTCAATGATACACGTTATATATGTCATAAACTGAAGTAC    4000
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GCTAGGACCAAGTGTTGACGACTGTGAGGTTCAGATTATGGCT-----AGTTTTACGAAC    3241
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        CCATCCCACTCAATCGACA----ATAAG----------AGTTCATACCCAACGA----AC    4042
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     TACGAGAAACGAATGGCCAGCATCTGAGAACTCAGTCAAGTCTATGACCTACGAAGGACA    3301
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        GGTTTCCCCCGTTTTATCACTCCGGAGTTGCTCACTCGAGGACGAGTTCGATTTGTTTGC    4102
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GACACCCCCCATCATAGCACTTGACAACAAAT-----GTCGTCCAGAAC-----------    3345
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        TTGATGGTAGTCATTACTCCCGTTGCCTCTCTCAAAGTTTGTTTGTTGTACTTGATACTC    4162
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     ----------------------TACATAGATGAGAGATGGTGTCATTTGCATGACAACT    3382
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        TATCCCCCCCAACCCGTCGCACCCTACAAAGCTTCGTTTCATTTCATTTCATTGCCCAGC    4222
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GTTGCGAAGCAACAGCAAGTCCAAAA--GAGAAGATTTGGAAGGC---AGAACGCGAAGA    3437
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        CCCCTCCAATGTGCCGCTTTTTATTGAGTCAATCCTGTTCTTTGTCTTGATCCCCATCAA    4282
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     CATGAAGAAGCTGACGCGGTACGTGTACCTGACTCC------GGGAGTGGTGGCCAGTCA    3491
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        AAGTAAACATAATTAGAAGTGAGATGGTTCAGATTTCACCTTCACAGTGTACATGAACAA    4342
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GTG-----AGACTTAGA-------GCGGTGGTGTGACCCG---------------GACCA    3524
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        TCGTCTCACCTTCTCCTCTCTCCACCCAATCCTCCACCACTACCAAACCTCGCTAACTTT    4402
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     TGTGATGATGTCTTTGTATCTGGCGTGAAGGATACACGAGGAGGAGAAGTGATAGTACTG    3584
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        CCAT-------AGGAGCAGCACCTCCACCACTAAAATCAGCAGCCTTCACATA------C    4449
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GAGTCAGTGCTAACGGTAGTGTTGTCATGGGAATAATGAGCAGCCAACGCATTCTGACGA    3644
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        AACACATTATTACACCGAATCAACACCTCCCCCAACGTACCCGCAAACTGTCCATCAATG    4509
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     AATGAAGAATAACGGGACATGAAAGAGTC-------------GCGAAGATGGCGAAAATG    3691
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        AACTCCTCCGTCTCCTCCAATTGTAAATTCATGTAGGCATCCGAGGAGGCGAGCTTGCCT    4569
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     AGCACTGTCTTTCATCTCTTTTG--------TGAAAAGATCCGCGGGATTGATCTTGCCA    3743
bd_1474              ------------------------------------------------------------    1357
```

```
PDS1_Extended        TGGTATTCTTGGCCCCAC------TTGA----------GACGGACGCGGACGTGTTTGCC    4613
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GAAACATGTTTCACTTCAATGAGTTTGTGCTGCACCCACTCGCGCACAGCGTTTTCTTTC    3803
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        TGTTAAGTCGGC---------TAGGA------------AGGGTTTGGGGTTGACGATTAG    4652
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     AGTTCCATGTGTCGAATAGCTTTGGAAGTTGTGTTGTGGGTCCATGAGACGCAAGATTCA    3863
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        TTGTGACTGTTGATGAGAGTTTGGTGGAGGAGGAGGAGATTTATAGGTGAGAAGGTGGGC    4712
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     TTGTAATTCCAGATAGTAGTG--------------------------------------    3884
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        AGAGGCATCAGACGAAGGCATCGGCCTCACCATTGAACGAGCAGATAATAGATTAATCAA    4772
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     -GCACCGTCAGCGTCAGAAATGGGAACGCCAATAGAACAGAAACCATTGGCGATGTTGCG    3943
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        TCCCTTCTAGATGGGCTTACCATTGTATTCGGGGCCGTTAGCGTTCGAAAGAGTGAAGCG    4832
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     AATAGCCAAAGTGAGTTTCCCAGCGACATTGGTG--------GCTCGAATCTCTGCTTCA    3995
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        GG----TAAAGGCACCGACCGTCTTGTCGTCGTCGTCGTGGTGAACCAAATGGTCCCGCC    4888
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     CAAGAACTTAGTGAAGTTCTGTCTTGGCGTTCTGATGCCCAT----GAAATCGGTCCGCC    4051
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        GTTTCTGAAAACAACAAT---AGTGACAATATGCTGTTAAACGAGAGGTTGGGAGTAACC    4945
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     AGATCGAAACACAATAGCACCACTCATACTACGAAGCTTGAAGAGGGGAAGGAGCGTT-C    4110
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        AAGCCAAGCCCGCAACGAAGGCAACAGGAGACCAACAA----------------------    4983
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     CTTCATGTACCGCGTTGCCGATTTGACTACCCCAACATGCATCACTGTAAGTAGTCAGTT    4170
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        --ACAACCAGAACTGATCAAAACAGCCCAAACACCCAACAATGAACACGCTAATTGCACG    5041
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     TGCTTTCCATACCTGATCTTGGTGGTGTTGCATCACAATAAGCTTCA-----GTGTCCGA    4225
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        TGGCGCCGAGAAGTCAAGGTGTTT--GTTGCGCACTAGCATCAAAT--------------    5085
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GGAATCAGGAAAGTGAAGATATGTATGGAGAGGATTACGGTTTTCTGACGTGAAGGTAAT    4285
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        --------CAAGTGCACTCTCGACGTCGGCGGG------------TGCATCCT---CTCC    5122
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GCCATAGTCATGGGTAGAATGGATATAGTGCAGTGCATGAAGGGCTGCTTTCATGTGACC    4345
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        AAGAAGTGATGTG---TCAGCACGTAATCAGCCGCACATGATGCATGCTACAACATCATC    5179
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     AGGAGAAGGCTTGTTGTTGTACGATGATAAGAAGGAGTGGACTGGAGCAATGTCAGGACG    4405
```

```
bd_1474              ------------------------------------------------------------   1357


PDS1_Extended        GGCGTCTGCTGGCAACTGCTACACTAACATCCAACCATTCTCGGC--------CAATACC    5231
PDS1                 ------------------------------------------------------------   2171
bd_1474_Extended     AGTGCATTGTGCCAACCA--TCCAATAGATCCAACAAGACTCATATATGCCTCCGAGCAT    4463
bd_1474              ------------------------------------------------------------   1357


PDS1_Extended        TCCAGTTGGACCCGCAACT---TCTCATC---GGAAGCACCCTCACAACAACAACAACAA    5285
PDS1                 ------------------------------------------------------------   2171
bd_1474_Extended     CGTTGTTGGGCAGTCGAATGCGTTTCATTGTTGGAGTCAGCGATTGCATCAATAGGAATG    4523
bd_1474              ------------------------------------------------------------   1357


PDS1_Extended        CAACAACGTCCCAGCGCCATCGAAAAAGCCAACCCCCCCCTCCTAACACCCGGACACGCC    5345
PDS1                 ------------------------------------------------------------   2171
bd_1474_Extended     CCAGATCGATACGGAGTGATGGTAGGAGAAG---GAGTCTTAGAATCA-ACTTGGAATCG    4579
bd_1474              ------------------------------------------------------------   1357


PDS1_Extended        ATCCGCATCCTTCAAATCCACGGAGCCAACCAAGCACAACTCATCCGTCGTTCAGACTTT    5405
PDS1                 ------------------------------------------------------------   2171
bd_1474_Extended     TTCCGCAAGATTT--------GCTGCAAAACCAGCCTGATTCATGTGCACTTTAACTGTT    4631
bd_1474              ------------------------------------------------------------   1357


PDS1_Extended        GTCAAACTATG------TGAATCATCCCGTCCTGGAAAGAAACGGGATGCCAAGGTGATT    5459
PDS1                 ------------------------------------------------------------   2171
bd_1474_Extended     GATGGAGTAAGACTCCAAGAGAAGTGTGTGCCCAAAAACCATTCAACAACGCCCATGAAG    4691
bd_1474              ------------------------------------------------------------   1357


PDS1_Extended        GCTACGGCTTTGAGGGAGTTCAAACGCAATAACAAGTTTGTGCTTCATAGGGAGGGAGCG    5519
PDS1                 ------------------------------------------------------------   2171
bd_1474_Extended     TCAACGGCGATGAGGTTTGACAGAATTTCCTGGAACTTTGTC---TCTACAGAGTCGTCG    4748
bd_1474              ------------------------------------------------------------   1357


PDS1_Extended        AGTGCCGCTGTGGTGGGGATGATGAGGTCCAGTATTCCGAATTATAAGGTGGTTTATGGA    5579
PDS1                 ------------------------------------------------------------   2171
bd_1474_Extended     GGTG--------------AGTAATACACGAAGTCATCAACATACAAACCAAGTGTTAGTG    4794
bd_1474              ------------------------------------------------------------   1357


PDS1_Extended        AAGCCTAGGGTGGAGGCTGCGGTGTTTGTGGCGGAGCAGGTTTTGGATGAGAGTACGGGG    5639
PDS1                 ------------------------------------------------------------   2171
bd_1474_Extended     GTGCCGTGGACGG---------------TGGTACAGATGGATTG---TTAGGGTC----G    4832
bd_1474              ------------------------------------------------------------   1357


PDS1_Extended        TTGTACTTTGCAGTGAAGATTGAGGATGTGGATGGTGTGTTGAGTGAGTTGCAGCAGGGG    5699
PDS1                 ------------------------------------------------------------   2171
bd_1474_Extended     ATGATGTGTCGAGTAAAGAGGCATGGATCCGA-------TGGATTGGGTTGCAAACCCAT    4885
bd_1474              ------------------------------------------------------------   1357


PDS1_Extended        TTGAGTGAGTTGGAGGAACGGGGTATGAACGTTAGACTTGCAGCTCCAAATGAAGAAGAG    5759
PDS1                 ------------------------------------------------------------   2171
bd_1474_Extended     CGACA---TGAGGATGGAGCGGATCTTATCGTACCAGTGACGGGGGCT-ACGACGAAGAC    4941
bd_1474              ------------------------------------------------------------   1357


PDS1_Extended        GTCGAAGAGAGTGGTGATGAGTGCAAGATGCTACAAG----------------------    5796
PDS1                 ------------------------------------------------------------   2171
bd_1474_Extended     CGTACAATGTCTTCTTGAGAAGCCAGTATTCGCCAGGCTTAGCGTCGGGATCACCGGATG    5001
bd_1474              ------------------------------------------------------------   1357


PDS1_Extended        -----------ATG-------CTCAGCGTGTCACGGAAGGATTGATGAAGCTATTGGTA-    5837
PDS1                 ------------------------------------------------------------   2171
```

```
bd_1474_Extended      GAGGGCGGACGATGGTGACTTCGTCAGGTGGCAAAATGCCATTGCAGAAGGCATTTTTGC    5061
bd_1474               ------------------------------------------------------------    1357


PDS1_Extended         --------------AAGCGAAGATCTCGTCCAGAGAACAAGATGAAGAAGAGAGCAAAGC    5883
PDS1                  ------------------------------------------------------------    2171
bd_1474_Extended      AATCACCTTGCTTGAGGACACGACATTGTTCGATAGCCAAACTG-ACAAGGTACCGGAGG    5120
bd_1474               ------------------------------------------------------------    1357


PDS1_Extended         GAGGGTATCTCAAACTCCTTCAACTGAAC-----GATGGCCCGGAAGACAGCACCTTGAA    5938
PDS1                  ------------------------------------------------------------    2171
bd_1474_Extended      GAGTCTTGA-CGAAGAACTGGTGCGTAACGTTGAGATTTCGACCAAGAGCGATCTTCGAG    5179
bd_1474               ------------------------------------------------------------    1357


PDS1_Extended         ACTGGCAACACAAATCGCTCTTGCTATTGGGGGATCTGCTTCTGCGAGGAAGAACATTAT    5998
PDS1                  ------------------------------------------------------------    2171
bd_1474_Extended      GTTACCAAGGACAACGATGCGTGATTTTGC----------ACGAAGCGGCATGAGATTTT    5229
bd_1474               ------------------------------------------------------------    1357


PDS1_Extended         TGCTCCGTATGCTGATGC-----CTTTTGGAGTGGAGAGGTTGACGAGTCTATTCTTCAG    6053
PDS1                  ------------------------------------------------------------    2171
bd_1474_Extended      CATCCTTTTTGATGGTAAGAACACACATGGTAGGTATGGCTTTGGGAGCACCTTTCTCAC    5289
bd_1474               ------------------------------------------------------------    1357


PDS1_Extended         TTGGTTGCGGATGCTGAG----GCGAAGGAGTTGGCGGAGAAGGAGAGTGTCGAGGCTGC    6109
PDS1                  ------------------------------------------------------------    2171
bd_1474_Extended      GCAATGCCCGATATTAACCGACAGTGAGGCGTTGGAAGGTACCCATACTTTCAATAGAAC    5349
bd_1474               ------------------------------------------------------------    1357


PDS1_Extended         CGCCGCCGCAGCTGCTGCTGCAGAAGAGGAGGAGGCATCTGAAGCGGGGGATGGTGAAGT    6169
PDS1                  ------------------------------------------------------------    2171
bd_1474_Extended      CCT-------TTTCTTCATAGTAACTTTGAAGCCATATTT--------------------    5382
bd_1474               ------------------------------------------------------------    1357


PDS1_Extended         ACACGATGGCAGTGAAAATGATTCAGAACAATCAGAAGAGTCAACGGAGGAAACTAAGTC    6229
PDS1                  ------------------------------------------------------------    2171
bd_1474_Extended      -CTCGATCGGGATGGTTGTCGGCGAGAGCACGAAGGAGAGTGGCAGGA---C----AATC    5434
bd_1474               ------------------------------------------------------------    1357


PDS1_Extended         TTAAGGTAATATACTGTGTAAGGTAGTACTACACCTGTTGTTATCCGATGGGGTCCGGTT    6289
PDS1                  ------------------------------------------------------------    2171
bd_1474_Extended      TTGATG-------------GAGGTTGATGGCACTGACGTGATGTGCAACTGGATCAAACG    5481
bd_1474               ------------------------------------------------------------    1357


PDS1_Extended         GCTTGGCTGTCGAGGTAGGGGGGGAGAGGTGTCAGCCTCGTCTACTGGGTGGTTGTAACAC    6349
PDS1                  ------------------------------------------------------------    2171
bd_1474_Extended      TGCTACCGGTAGAGGGAATGCAAAGG--AATGA-------TGAATGATGACCCGGAAGAA    5532
bd_1474               ------------------------------------------------------------    1357


PDS1_Extended         GCCTTCCTCACGCTCCGCGCTCCCCCCCCCACCAACCACAATCCAAAAGCAAACGCCGCCA    6409
PDS1                  ------------------------------------------------------------    2171
bd_1474_Extended      GAATGCCTTGAACACATAGATCAGTCCAATTGTATGGCAGGTCAGGAAGAGCAACCCCCC    5592
bd_1474               ------------------------------------------------------------    1357


PDS1_Extended         CTGCATCGTCCTACCCTGTTGTGC----TTCCCTGACAGCGCCACTCTTAAATTGACCCT    6465
PDS1                  ------------------------------------------------------------    2171
bd_1474_Extended      ATTCCTCTTTGCGTGCATTTGGATGGCGTTTATACACAAACCGGTATCCATTGTCGGACT    5652
bd_1474               ------------------------------------------------------------    1357


PDS1_Extended         TACACTGCTGATATATGCCATCGGTGCCGTATTGATTGAACGTCGTCCTACGATTTGCAA    6525
```

```
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    TACCCAAGTAACCTTTGTAGTATTGTCCCTCATGCTCATAAGTGATCTTACTGTTCAGTT    5712
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       ACATTAAT---CAGAGAACATCAGCTCACACATCATCGGCATCGGTGCACGTCAACAACC    6582
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    GAAGAAACGGTGGAAGGAGAGTGGAGGAGTCATCATCTGACGAGGGAGCTG---------    5763
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       ACCGCAATCATGCTTTCCGAAACAGGCAAAACAGTCCTAGAACAAAGTCCCTACACCCTC    6642
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    --------GAGGCGTTGGGATTAAGTTGGGCATGTCCGAGAGGGAGATATAGGCAGTCGT    5815
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       AACATGGTGGAGTACGATTTCGAGAACTATGCTCTTACCCAAGGTAAGAAATCCGCCTAT    6702
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    AGAGTCAT-----CAAATTGGATGAGATAGTCACTGTGAGAAGGAGAGGAC---------    5861
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       GACTTGCCCAAGTGATCTACGATGCAGTAGTTCAGCAGAGTAATTGTAA-ATGATCCTAA    6761
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    GGGAGGGGGATGTCCATGACAGTGCCGGATTTGAGGA---TGTTTGTGTGGGGATCCACG    5918
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       CCTGTTAATCACTGCCATGAACGAATGACGTCTTGCGTCTTCACCCAAGGTATTCCCGAA    6821
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    CGTTCTACGCGAGTGCCTGGAGGATAGGGTTCG-TCGAAGGAAGGAGTGGTGTCACGATA    5977
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       ACACTCGACCCTCTCCCCTACCGCAAGGCTGGCTCCG----TCTTCGCCACGGGCGAAGA    6877
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    TAAGGAGCAGAACAAGCCCCCATCATAGGTGATTGTAGGATAAACCGAAACTGGCAAACG    6037
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       GGCCGAGCCGGTATTAACCTCCGCCAAGCACGATCCTAGGTTACGTACCGTGGTGTGTCG    6937
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    AAACGGATCAA--TACGATAGCTGTCAGGCTCATAGTACTGTTTGTTGCGTGGATTGTAG    6095
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       TCACTGGTTGCGTGATTTGTGTATGAAGG---GAGCGGCGTGTGAGTTTTTGCATCAATA    6994
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    ACCCGAATAGCATTGGATGTAGTGGAACGACCGATGACGATACCATCCATCGTATGAGCC    6155
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       TGATTTATCCAA---------GATGCCACTGTGTCGCCATG--------------------    7026
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    TGATTCTTAGAACGAACACAAGAGCCGTCTTTGTTGTGGTGGAAGTAACAGACGGAGAAG    6215
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       ---GCGATCGGTGTAAAGTGAAGGATTGT----CCATTTAGACACATCAACGAAGCGGAT    7079
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    ATAGGGATCCAAGTACGCTGATCAGCTTTCTCACCGTGTACGAGCATGAAGGGTGATGCT    6275
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       CGGTTGGAGTGTGTCTTTTACAGTCAGGGATTCTGCATTCACGGTCCATTCTGCAGGTAT    7139
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    AATTTACCGTGGAACTTT-------------CCTGGAATCATG-TTCATCATGCG---AG    6318
bd_1474             ------------------------------------------------------------    1357
```

```
PDS1_Extended        AGGCACGTGAGGAGGGAGAGGGCCGACTTGCCAGTGGTGGCTGATTTTACATTGGGGTTG    7199
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     CGGCATGAGAGAT-GGAGTGAAACCAAAAGCGACGCGGCATTTGTTTTTCAGTGAGGTAA    6377
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        AGTCAAATGCAAGCTGGGAAGGATGGAGTGTCGGCTGTACGACGACCTGCTCCGAAGCCA    7259
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GCA----CGAGACATGTGAACCATCACC--------TTCCAATGGCTTTCAACCAAACCG    6425
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        AATGAGTTCTT-------CA---AAGTTAGTTTGTGTAAGCACTTTTTGAGTG--GGGAG    7307
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     TTGGAGGACTGGCGCCCTGCAGGAGCGGCAATAATGTTGGAATCTTTGTCTTGGAGGTAA    6485
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        TGTCCCTTTGG-AGATGGATGTCACTTTGCTCATG--------------------------    7341
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GCTCGGATGGCAGTACCGAAGAGCTTTTCGTCACAGTCACAGCGAAAACATTTGGCGAAA    6545
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        -GGGAGGCAGAGTTGAGGAGGCCAGGGCAAGAACCGACGAAGACTGTGGGGGAGGATGGG    7400
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     CGGCCGGCTTCACTCCTGAATTGGTTGAAAGCAGCGATGATGTCTTGTGGAGAGAGCGAG    6605
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        GAAATTGCGGAGAAC---------ATGTTTGGGACGCAGGAGACGACTACGGTGGATTAT    7451
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     CGAAGACCAAAGACCCAAGTGTATCTGGTTGCTCCATCTGCGAAGACAAGAGCATA----    6661
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        TATCAGGGTGGAGCTGCCGGTGGGGGTAAGCCAACG------CCCATTTTGGAACCCGAG    7505
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     ---------CATCGTGCCACCAAGGGCAACCCCATCACCAAAACCAATGTCAACATGAAC    6712
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        AATGCATCATTCTTTATTCTGCGGGCGGCGACCTATCGTG-------ATTGTGCCATCTC    7558
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GATGTCCAAATACCTG---TAACGAGTACGATCGATTGGGGTACCACGTTTTGCCTTTGG    6769
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        TACTGTTCGTAATGAATGGATGGTACAACGAAAGCATGCAGAGGCAATCAAC---AAAGC    7615
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     TATAGTAGCGTACGATCCAAGAGCTAATGGAAACTCACCAGAATCCGTCCACACTCCATC    6829
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        TCATGAGGGTGGGAGACAAGTCATGTTCTTCTTTACAGTTGGGGATTCCAACCACATTCA    7675
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GCGAGAAGTTTGGAGAAGGTGCTTGTAGTTCTTGAAACGACGGC-----AACCAGTGATA    6884
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        GGGGGCAGCACTGCTGACTTCCTCCGCATCGTAT------GTTGAATCGGCAGATCACAA    7729
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     CGGTGGAGTTCCTCTGCTGACCAATGGGTTTTGGTGTCACTCCCATTGGGCGTGTCACAT    6944
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        ACAACCAAAGAACGGAGCTCCTACAAACGCCAACGCAAACAAAGATGGAGCATTCTGCTA    7789
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GGACGGATGGGA---------------GGAGGAGACGTACCGGGATGGTGGATGAGGCGG    6989
bd_1474              ------------------------------------------------------------    1357
```

```
PDS1_Extended        CCGATTCAA---GTGTGAATGGT---ATCGTACTTGTGAGCTGCCCATCTCCACTGCATT    7843
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     AGGATGTCGGTGGTGGAAATGGTGGAAAGGAGCCGAGGTGCTGCGGATACGGGCTTGTCA    7049
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        GGAAGCTGCTCCCGATCTCTTGCTTCCCACATCAACTACACAATTCTGTCAAGATATGAA    7903
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     ACAAGTGGGATGC-----GTTGTTTGATCTCATCAGGAGCCAAGTCAGACATTGAA----    7100
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        GTCTAAAACTGGGGAAGCAGTGATGAAGGCAATTTGGAACTCGCCTCTTTGTACGTTGTA    7963
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     --------CGAGGCA------------CT-GAGTTGGAAGGAGGCTGGGTGTCAGGTGGA    7139
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        CGAGTCATGGAATGGAGGGGACTACGATGAAATGGGCGGGAAGGACTCACGTCCTCCTCC    8023
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     A-TGACATGTAGTACCGGCGGAGATGACGAAGTGGCTGCATTGTCCACA--------ATG    7190
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        GCCAGTGGGAGATGCAATATTGACAGACTTTCGTTGTCCTTTGCCGGAGGAGGT---GTC    8080
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GCAACCGGGCGAGCAGAGA-CCCGAGGCTGAGCGTAATCTAACTCCGAAGCGATGACGTT    7249
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        GTGGCCTACGCTTCCA------GGTCCGGGGTTCATATTTGG-GTGCAACTCGCAGACTA    8133
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GGTGCCTAACGGTTCATATAAAAGATGACAGTCCAAGGTTGTGTTGACAGTGAGAATGAA    7309
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        TGGATGAATGTTTAGGAAGAGGATTGTTTGGTCTACCAGCTCATATGAAGGTGAGACTTT    8193
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     TGAAGGAAAGTAGACATACATACCCCCCATGTCAGCGTCACCAATGAAACCGCAGCCTTT    7369
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        GTTGGGA----GGGCACAATGTAGTTTGAGACTATTTCCTTAAT-GCTCACTTGGACTCG    8248
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     CTGGAGTAAGTGAGACCGGAGACTGTAGAGAGGATTGCGAAGAGCCGGAACATGCAGACA    7429
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        TTTGCAAC-----------TATTCCTCTAGATTGCTGCTGCAGGTATCCGACCAG-----    8292
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GTTCCGAACCATAATCTTCTTTCCATTGAGAGAGAAGACGGCAGTACCCACCCCACAGAC    7489
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        --GCTCTTCCATCTT-TCTCTACAACGTCAGTGAACGTCTCATCTTTGGTATCTTTGAAG    8349
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     TGGAGCCAACGACTTGTTTCCCATGCGCACCTGCAG-GTGAGTGATGGGTTTGTAGGAGA    7548
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        CTCTTACACCGGCTATCATGAACATGGAACCAAAGGCATTCTCAAAGAATCCAAAAGCTC    8409
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     CAAATACCGACTTGTCAGGGAACATGTGA-----------TCAGTCGCACCAGAGTCAG    7596
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        AAACCTCACCTTTCCCTGTACAAATCCGTGTACGAATCAGTTTGGAGTGTCCTCCCGTCC    8469
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     CAACCAAAAGTGCTGAGGATG----GAGTGGAGGTGGTGATGTGGGCTGGTGTAGTGTCC    7652
bd_1474              ------------------------------------------------------------    1357
```

```
PDS1_Extended        ATGAC---G--------------ACGATCCAGCGTTGA-----------ATGATGTGTTG    8501
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GAAATGGAGCCCTCCAATAAGAGACGGACCGACGTAGGGAGCCGAAACACAGTAGTGGTG    7712
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        AGAGCTCGTGTTGGCGGAAGAATTGGACCACTGACGTTTGCACAGTCGGAGGCATTGGCA    8561
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GAGGACGATGCCGATGGAGGAGATGGACGAATGGGAGAAGGACAAGAAGGGGCAGTGGAG    7772
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        AGTTTGATTGCCAACCAATGTGGGGCATTGAGTTACATGACAGAGTATCGCCGTAGTATT    8621
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GAGACAATGTACTTGGAACTGGAGTCGTTATGTTTAGCGACAACA--GAAGCATAAGAGG    7830
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        GAAACGAGACAGTTGAACGTTGCA---GCTCCTCCAATTGCATTGCCCCCTCGTAAAATT    8678
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     CAGACGAAGTCGTATGATCTATAACAGCCTCATCGCCATCCCAAGCAAAGTCATTGAACT    7890
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        GGAACGGCTTAAGGGGGACAATCTTGACTGTTTAAAAGTACACATGAATTGATATTTTAA    8738
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     CCTCAGGCGTCGAGTACCCCCTATCGTCC-------GCACCAGATGACG-----CGGTAA    7938
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        CGATACTTCTTGTTCTCCTCAGTATTCACAATATATGCCATGCTTCGAAAGAGCTTATAC    8798
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     CCTGACTGACAG-----CTGGCTGTGTGCTAGACCCCCCCTCATTCAGCTGAGACACAGG    7993
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        A-TAAGGGAATGTCCACACTCGTCAGTGCCCATCCATGAGATTGAAGCCG----------    8847
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     AGAAGAGGGAGGAGGTGAAGGGTCAGTGACAGTAGGTTGGACAGCAGCAGGTTGAGCGGA    8053
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        GCAACACTTTTAAATTCATCTCGGCAAAAACCTGCATCGTGCTGTTGCTGTCACTGAAAT    8907
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GGGAAGAGACGACTTGCCTTCGCGAGTGACGGTGA-----GATGTGCCTTTTTCAGCAAC    8108
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        GGGTCTGGAGTACTTTCTGTCGTGCTACCTCAGCTTTGTTGGTGTTTTCGTTTAGCTCCA    8967
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     TGGCACTTTGTGATGTTGTGACAGCGACGTCGCATGCA--GATGCAGTTGGACTTCTTGA    8166
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        GCAGCGAGGTAAGCTCTT------------CGGGTACAATAACTGGGGAATGATATTGGA    9015
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GAAAGAAGGCAAGAACGTCCTCCTGCTTGAGGGCACCAATCCATTCCCACGGAGTCTTGA    8226
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        AAAGAG----AACGGAGCGTGTGGTTGGAGAAGAGGGTGCTGTTGATGCTTGACGGGTCG    9071
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     AGGTTGAAGGGGCCTGGCTACTGGAAGGAGTGGTGGAAGCTGCTGCAGCTGCAGGAGCAG    8286
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        CAGAGCACTCGGCAGAACTCCCTCACACCCTCGGGAGTTATGTTCTCGCTTGTGACGTCG    9131
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GAGCAGGAGTAGCCGCGGCAGATCGAGTCCCAGGAGCATCGG--------TCGAACAAGG    8338
```

```
bd_1474              ------------------------------------------------------------   1357


PDS1_Extended        AGTATTTCTAATGTTGTGTTACCAACCAAGCTAATTGCCAATCCCCTCAATACTTCATTG   9191
PDS1                 ------------------------------------------------------------   2171
bd_1474_Extended     AGCCCGATCAAACTTCTGGACAAGATCAGACACCTT------AGACATTGTCATGGAGTC   8392
bd_1474              ------------------------------------------------------------   1357


PDS1_Extended        GTAACCGCGTTGCAACCGAGATCGACAACTTCTAGCAATGTGTGGTTACTTTGGAA----   9247
PDS1                 ------------------------------------------------------------   2171
bd_1474_Extended     G---AAGGAGAGCTCACCAGAGCGAAA---ACGGTCAATGAGTGGCTCGTATTCAGCACG   8446
bd_1474              ------------------------------------------------------------   1357


PDS1_Extended        --GACCGTCGACAACAACTGCAATCCAGCTGCAGTAAGACTACGATTGTCGGATAA----   9301
PDS1                 ------------------------------------------------------------   2171
bd_1474_Extended     AAGACCCATTACAAAGAACATGACCTGAAGTTTGGAAGACACCAATTGTCCCGAGGAGGA   8506
bd_1474              ------------------------------------------------------------   1357


PDS1_Extended        ---ATCGAGA-----CGCCTCAACTTGGAGTTGGTTATTAATCATTTGTCAATGCAAGCA   9353
PDS1                 ------------------------------------------------------------   2171
bd_1474_Extended     AGTACGGAGGAACATCCCCTCCAACCGAGATTGGTAGGAAGTGACTGGTTCCGACACCCC   8566
bd_1474              ------------------------------------------------------------   1357


PDS1_Extended        TAGCTTCGTCTTCAAACTGATTGCTTGATACCACAAGAGTCTCCATTGCTGTGTTTGGAT   9413
PDS1                 ------------------------------------------------------------   2171
bd_1474_Extended     ----------TTGGCGGTTGTTGGGTAAATCACGGAGAATCTCAAATGTAACCATATCAT   8616
bd_1474              ------------------------------------------------------------   1357


PDS1_Extended        TTTGCAGTATGGAGGACCGAAAGAAGCGTGACCATCCTCGTCCTGTCGCTCTCCCGTTGT   9473
PDS1                 ------------------------------------------------------------   2171
bd_1474_Extended     CAC--------ACGGAGAGAAATGGTTGT-----TGATTGCATCAACCATCTCAAATCCT   8663
bd_1474              ------------------------------------------------------------   1357


PDS1_Extended        TACCGAGAGTTAAAGTCTTCAGGTTTGTGTTTGCAGCGAGCCCATTGGCCAAAGACAGAA   9533
PDS1                 ------------------------------------------------------------   2171
bd_1474_Extended     TTGCCATCATACAAATC---AC-----------------CCTTGTTCGCAAAGAGCAACT   8703
bd_1474              ------------------------------------------------------------   1357


PDS1_Extended        GTGATTCGTCGTCAAGACCACAGCATGAGGCGTTCAATTCTTTCAACGCTGAATTGGGCA   9593
PDS1                 ------------------------------------------------------------   2171
bd_1474_Extended     GTGCCTCGGAGCCATCGACA---------------------------------------   8723
bd_1474              ------------------------------------------------------------   1357


PDS1_Extended        TAACTTCTGAAAGCGCCTGCCAACCAATACTGGTTATATCAGGATTCAAGCTCATAGTCA   9653
PDS1                 ------------------------------------------------------------   2171
bd_1474_Extended     -------------------------------------AGGCAGGTGTGTACCTTCAACTCC   8747
bd_1474              ------------------------------------------------------------   1357


PDS1_Extended        ATGTTTTCAATGTGAAATTGCTGCTCAACCCATGTGATAATGCG----------GCAATA   9703
PDS1                 ------------------------------------------------------------   2171
bd_1474_Extended     AGCTGACGACTCTGAGCAGCGTTGTCAACAGTGGTGACCAGAGGGTCACTTGAGGAGACA   8807
bd_1474              ------------------------------------------------------------   1357


PDS1_Extended        TCTTCATCCGTCAGCCCAGACGAGATGATGATCAGGTGCTCCACTGTGGAAGTACTCAAA   9763
PDS1                 ------------------------------------------------------------   2171
bd_1474_Extended     TTGTAGTAGCTGCGGAGGAAGAAACGAGTGCTGAAGATCCAAGAGGAGTAAGAAG-----   8862
bd_1474              ------------------------------------------------------------   1357


PDS1_Extended        GCAGATGCAAGTGACTGCCATGTTC------TCACCGTTGAGAACAAGCAATTCGCG---   9814
PDS1                 ------------------------------------------------------------   2171
```

```
bd_1474_Extended      AGAGATCCGAAAGTTTCGGAAGCTCATACTTACGAGCTTGAATGGCATCGTAGAGGGAGA    8922
bd_1474               ------------------------------------------------------------    1357


PDS1_Extended         --AGGTTCAATGTTTGCAGTGTCGAGTTGCCACTCAAACCACCAGAGAGCATTGATATTC    9872
PDS1                  ------------------------------------------------------------    2171
bd_1474_Extended      GCTGGTGTGATGGAGACTGAATCAAATTGGTGGTAGAG-ACCTCGAGAGCAAGACCCTCC    8981
bd_1474               ------------------------------------------------------------    1357


PDS1_Extended         CATCGTCATTCATCTC--------A---------------TTATTGCTGAGGTCAAGCCG    9909
PDS1                  ------------------------------------------------------------    2171
bd_1474_Extended      AATTGGTCCTCCTCAGGGACACACAATGACAACGAAGCGGGGATGGTCGAAGCCACAGTG    9041
bd_1474               ------------------------------------------------------------    1357


PDS1_Extended         TTTCAATTGGGAATTTGAGCTTTTCAGCATGTTAGATACTCCCGTCATGAGTACACAGTA    9969
PDS1                  ------------------------------------------------------------    2171
bd_1474_Extended      CTAGAAGCCGAAGGTGGAGCCGGCGAGGGTGCAGGATGACGCTTAGAAGGG---------    9092
bd_1474               ------------------------------------------------------------    1357


PDS1_Extended         TGAACGATTTGTGAGCGAGCTTGCATCAGAGAGTACTAATGTTTCAAGAGCCGAAGCTGT    10029
PDS1                  ------------------------------------------------------------    2171
bd_1474_Extended      GGGAGGATAGAGGCAAGAACATCAATG-------TCGTCCACCTCATCATCCGAATCGGA    9145
bd_1474               ------------------------------------------------------------    1357


PDS1_Extended         CAATCCATTAGACAGGGCACCAATC-------------CCT--------C--TACCATCT    10066
PDS1                  ------------------------------------------------------------    2171
bd_1474_Extended      CCCCCCATGATCGGCGGCAAGACTCTGCGATGCCAAATGCTGAGCCGCGCGACGCCGTGT    9205
bd_1474               ------------------------------------------------------------    1357


PDS1_Extended         GAAAGGCTATCTTTGATTGTAAGCGTTCTCATATTGTTGTTACTGGAAAGAAGAGTCTCA    10126
PDS1                  ------------------------------------------------------------    2171
bd_1474_Extended      ACGTTGTGATTTCTTCTTCCGAGATTGATAAGATTGGTTGTTGTATAACGCCGATGGAGG    9265
bd_1474               ------------------------------------------------------------    1357


PDS1_Extended         ATTGCTTCGCAGCTTTTCTTTCCGACGTAGCTACCCATCACAGAGAGACTGGCTAGATTG    10186
PDS1                  ------------------------------------------------------------    2171
bd_1474_Extended      CTCCA----GCTGAACCCGGGCCGCAGAGGACAAAGATGACACTGAAG-TGCGTCCACTG    9320
bd_1474               ------------------------------------------------------------    1357


PDS1_Extended         GGGTTGGCAGCCAATGCCTCTATTATAAGCCTGTTTGCTTCGTCGGTACGAACAATACGT    10246
PDS1                  ------------------------------------------------------------    2171
bd_1474_Extended      GAGGGGGTAG----------------ACCGAAGGGAGC-GGGCAGGCGAGAACAACAAGT    9363
bd_1474               ------------------------------------------------------------    1357


PDS1_Extended         ACCT-----CCCATTGATGTTGGTGAATATTCACACGTTTCAGGGATGTG--TTTCCTAG    10299
PDS1                  ------------------------------------------------------------    2171
bd_1474_Extended      GCTGAGGCACGAAGTCGTCATCAGGAGGATCCACGCCAAGCCGAGAAGGGGGCTGAGTGG    9423
bd_1474               ------------------------------------------------------------    1357


PDS1_Extended         GAGCATCGATGCAATACATGATGGATGAATGACAAGACATTCCTCCAAGTCGAGTTCGGT    10359
PDS1                  ------------------------------------------------------------    2171
bd_1474_Extended      AAGGAAG--AGGACGAGTTGATG-------CAGAAGCTCGCGAGAGAAGCCAAGTAGGAG    9474
bd_1474               ------------------------------------------------------------    1357


PDS1_Extended         AAGATGCCCGTCTGCGAACAGGGGCTCAAGTTCTGACCAATCAGCTCCGATAAAGCACCT    10419
PDS1                  ------------------------------------------------------------    2171
bd_1474_Extended      GAGACGCGATGGGGTTGCGTGAAGGTGTATTTCC-----CACAGCCACGGAAAAGGATG-    9528
bd_1474               ------------------------------------------------------------    1357


PDS1_Extended         GCATAACTGCAGCCTTATGATTGATCTGTTGCGA-----GATACCCCCT---CACAAAAG    10471
```

```
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    ACTGAGAGGGATGGTCATACTCCTCCTGCTGCAATTGGCGATGCAGGAGAGAGAGACGAC    9588
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       CTTCTAAACAGAAGTCTGAAGTGACCTGGTTCGTCTTCATCTCTATATCCACGGATGACC    10531
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    CCTCATGGAAGATGACAGGGCTACTCTGGAGGAGATCGGCGAGAAGGTCGACGGGTGCAC    9648
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       AACTGTTTCAG--------TTGCGTGTTCTCCCCGATATGTCTCCCAAGTTCCCCCAGTG    10583
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    CATTGTTTGCCGCAGAGAACTGGGCAATGGCCGCGTCCTGTGTCAGGGAAGTGGCGATCG    9708
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       CATCGTCTATTTCTTCTGTAAAACGTACCCTCAGTGAAGTAATGGTAG----------GA    10633
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    CGGTGTG-AGTGATGACACATCCCGTGTCCTCAAGGTAGGCAATGAGAGATAGGGCATCA    9767
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       TCGTTGGATCTAATCTTCGACATGATGTCTTCGCTATAATCTACGCACGCATAGTCAAGT    10693
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    GGGTCGGAGGGATCAGGGCAGAAGACGTGCCCGATAAGACCAGCACCGTCGGAGGGGGTG    9827
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       GACGAAACGAGATCGGTGAGCAAATCTTCATCATAAGATGCGTTGAGCCTCGCTATTTTG    10753
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    GA---------GGTGATGATAACAGCACCTTCCTCGGACGTCATGGTCGGCACGACGAGG    9878
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       CTCAA-----GAGATCTTCAACCGCGTTGAAGCTATCCATTGTACTAGTTCTAGACGTCA    10808
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    CCAGATCCAATGGGTAGGTAGGTAGAACGAGACTATAGAGTCG--AGGGTCTATAGGGAA    9936
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       AAG----------------------TGAGCCGTTGTGGTTGAGTATC----------GGTA    10837
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    AGGAAGGATAGCGAGATAACGGTGCTCGCCATGGAGGGGGAGCACCGTCTCGTCGGTATC    9996
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       ATCTTG-GGCCGCGAATCCGTGCAGGGGAAGGAATGGAAGGGGAACGTGTGTGAAGAAAA    10896
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    AAGTTCACGCGCTCAATTCGTGAGGTTCAGT----TTATGGCTA-GTTTTACGAACTACG    10051
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       GATACGTTGTTGTCTCCTCGTAATATCGTTTACCTGTAGAGGGGTCTATATGTAGAGACA    10956
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    AGAAACGAATGGCCAGCATCTGAGAACTCAGTCAAGTCTAGGACCTACGAAGGACAGACA    10111
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       ATTTATGGTCCTCGATGGGAAGGTAGATAAACAGGGGACAACTGATTTTCATTCTGTACA    11016
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    GGGTGGCGGCGGAGAGGGGA--------GAGGAGAGGAGAGATGTTGTCCCCTTA--ATG    10161
bd_1474             ------------------------------------------------------------    1357


PDS1_Extended       AACTGTTCCCGTTAGTAACACTTTTTATTTTAA-------------AGGGGTTAATTTAG    11063
PDS1                ------------------------------------------------------------    2171
bd_1474_Extended    AAACACACACTTTAGAAGGAAACTTTATTGTACCCATACAATAGAAATAAATTATATTAG    10221
bd_1474             ------------------------------------------------------------    1357
```

```
PDS1_Extended        TAGACAATTTATGGACACGATC-TCCCCACGGTGCTCGTGTCAGCAAACCG-CAACGAAC    11121
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GGGATTAAAAATGGACAATAAGATGGATGCCGAGTTCATAACCAACAAAAAAAATGGTAC    10281
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        CCTCT-CTT--CCGTTCCTACCCTCCCTCCTCTCCACGCCCACCACCACTCAACATGTCG    11178
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     CCTTCACCTTCACCCTGCTGCTCATCCTCCCTTATGCAAT-----CGACTCGAATCTTCA    10336
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        ACCCCCTCCCAGCTCCACA-----ACCCCAAAGCAGTATGGTTTCTCGGTGCACCAGTCT    11233
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     ATGGCCTACTTTCATCAGTAAGCATCGGCGACATCACATCTTCTCTACGTCGAG----GT    10392
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        CCAAAGCAATAGCTGCATCAACCGCTGCGATTTATGTGTTGGCGGAGATGAACAAATGGC    11293
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     CGAGCCCAGTCTTTG-ATGTCCGAGGAGGAAGTCAATGT------GGCTCGGAAGGTGGC    10445
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        ACGAGGCTCTTGTGTTCGGTGAGTGGTGCATGTGATCTGTGTTTAG--GGATGTGGCGTT    11351
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     AACAGTCAGTGGATTGGG---------GTAATTGATCTTCGTCCAGAAGCTGAACCCGAT    10496
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        GCTATTGAGCT---------------TGTAGTTCTTGAAGCCGTTGTTTTGTTGTCTTTC    11396
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     AGTATTACGTTGGTCCAAGACAATTCACAAGTTACTGAAAGAGCTGTATTCGTCTCTGAT    10556
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        TAATT-AA--TGTTCACAACTCTA-----------------------------GATACCT    11424
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     ACAAGTACCGTCATCACAGCTGTTGGTAGTGAAGCTGAAACGATTCAACAAACAGCAGCT    10616
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        CCAAGATATTTGACCAGGCTCAGTTCTACAGAATCTTTGTTTGCAATTTGACGTTTGCGT    11484
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GCCTTCTCCATTGGCAAGCTCAGTTCAACCATATATCTTGTGATATCCTATGACTTT-GA    10675
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        CGATTGGAGAACTCGTCTTCGGGTTGTTGGCTCTGTGTCC----TTTGATGCGTCGATTT    11540
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GCATGGGA--AGACAGTGTTGCATCGTACACTGGGTGGCAGCAAGTTGATGAATTTTGTG    10733
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        GAACGTGAGGTATGTTACGTTG---TAACTTATTGATATCCGCTGAATTGCATTGATTCT    11597
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GAT-GGAATGCGAGAACGGTGGAACGAACTTCATGCACAAATCAATTTGGTACTTGTTCT    10792
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        CTTGCTTGTTTCTTCAACTCATACGAAACGTATCCAC---TTCTTCACTCACAAGATGGG    11654
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     CGAGCCTTCTTCGGGAGCTGCAACATCGTCTTTTCTTTCCTGCCTCCCTTCAAGTACTGA    10852
bd_1474              ------------------------------------------------------------    1357


PDS1_Extended        TAGCAGAAAGTTTGGTGCATTCATAATCTATTCATCAGTCCTCTCAACGATATTTGAGTT    11714
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     TGATAAGGAGAGCGAGGAAACAAACAACGAAACCTGGGGCATTGCAAACACGGTGGCGTT    10912
bd_1474              ------------------------------------------------------------    1357
```

```
PDS1_Extended        GGTATT-CTTCAATATATTCTTTGACACAGAGCGATACTCCGGTCCATATCCACAACTGG    11773
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GGGTCTCACAGCACACACTGAATG--------GAAT-------------------ACCAC    10945
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        GGGCAGTCCTTGCAATGTATCACAAATTCGCACCTCGGCTGCATCCTAAGTTCTTTGGAG    11833
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GGGCAGAGATCACCTTGTTG----------AAAG---GC---TGGTGACATACTTTCAGC    10989
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        TGCTAGGATACGACTTCTCTGAGAAGTCACTCACGTATGGTTTATGTGCTCA-AGTGATT    11892
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     TGGGAGGAGAACAATACGC-------------CAATATCAATCCATTTCACAAGGTGGAC    11036
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        CTTTCGGGTGGTTTGAGCACGGCTATCCCTACAATCTTTGGCTTCATCTCGGGTATGCTT    11952
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GTTATTGGTGCTTGG-GATAGACAG-------------AAGGAGCATAGTGAATCAGATG    11082
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        AGTGTGAGCCTCAGTCAGCACGAGCTACCAGAAATCGTGTACACTGCTGCTGGAACTCTT    12012
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     AAAA-------CAAGCAAAGCAAGCTAGACGAC----TCCA---CGATGAAGGTAATCTT    11128
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        GG-------TAAAGCATTTGTGGACGATGCACCCGCAATCATGATGGCACGGACAGTACA    12065
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     ACGCCATGCCGAGTTGTTCAGTGACGTTGGTGCTGCAATCACCT--------CCAACAAA    11180
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        GCGT-GGAGGAAGGAATCAGCAAAGACGAGCTTCCACCGGTGCAAGGGGAGGAAGGGATG    12124
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     GAGTCATTCCAAGCTGTCATCCAAGAGGTGTTTGAATCTGTTGGTGGAAAAGGATGCTTG    11240
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        C-----CACGGCGGGA------------------------G--CTC--------------    12139
PDS1                 ------------------------------------------------------------    2171
bd_1474_Extended     ACTTTCCTCTGAGATACTATGGTCGTTGACTACTAATTAACCTCTCTGATATATCCCTTG    11300
bd_1474              ------------------------------------------------------------    1357

PDS1_Extended        --------CGGCAGCGGCACTTCCTGCTGTGCAGCCTC-----------------CT      12171
PDS1                 ------------------------------------------------------      2171
bd_1474_Extended     TCTATTCTTTGCAAACACACTTCAATCATCCCACGCTGACATGATCTTCGCAAAGCT      11357
bd_1474              ------------------------------------------------------      1357
```

Peptide Alignment (62.9% Sequence Identity)

```
PDS1      MIITNFILSTVLATSMAFQPHTPILSKPSFSNRVHRSPKIGSSNLV---MKDFP-KPNVE      56
PDS2      ---MKFLLPLLPAVAGAFSI-THLSQHPSLRMHQSLSTSLYSSSSSTSQRPRRPTPDRIR      56
              :*:*  : *.: **.  * : .:**:  :   * .: **.         *   .:.

PDS1      DTDNYRYAEAMSTSFKTSL---RVTNDSQKKKVAIIGGGLSGLSCAKYLSDAGHEPTVYE      113
PDS2      NTQNFKEAKELSQKFITDFQQLQKVGSGEPKRVAIFGGGLSGLSCAKYLSDAGHIPTLYE      116
          :*:*:: *: :* .* *.:   : ....: *:***:****************** **:**

PDS1      ARDVLGGKVSAWQDEDGDWIETGLHIFFGAYPNVMNMFAELGIHDRLQWKIHQMIFAMQE      173
PDS2      ARGVLGGKVSAWQDEDGDTVETGLHIFFGAYPNIHNLFDGLKIQDRLQWAPHRMTFAMQE      176
          **.*************** :*************: *:*  * *:*****  *:* *****
```

```
PDS1      LPGEFTTFDFIPGIPAPFNFGLAILMNQKMLTLGEKIQTAPPLLPMLIEGQSFIDAQDEL      233
PDS2      LPGQFTTFEFPAGVPAPLNMAAAILGNTEMLTLEEKIKMVPGLLPMLLEGQSFIDEQDEL      236
          ***:****:*  *:***:*:. *** * :**** ***: .* *****:******* ****

PDS1      SVTQFMRKYGMPERINEEVFIAMAKALDFIDPDKLSMTVVLTAMNRFLNESNGLQMAFLD      293
PDS2      SVLQFMRKYGMPERINEEIFIAMGKALDFIDPDLLSMTVVLTAMNRFINEADGSQTAFLD      296
          ** ***************:****.********* *************:**::* * ****

PDS1      GNQPDRWCTPTKEYVEARGGKVKLNSPIKEIVTNDDGTINHLLLRSGEKIVADEYVSAMP      353
PDS2      GNPPERLCQPMKESIEKKGGEVVCNSPVVEIQLNEESNVKSLKLANGTEITADYYVSAVP      356
          ** *:* * * ** :* :**:*  ***: **  *::..:: * * .* :*.** ****:*

PDS1      VDIVKRMLPTTWQTMPYFRQLDELEGIPVINLHMWFDRKLKAVDHLCFSRSPLLSVYADM      413
PDS2      VDVFKRLVPTQWSTMPYFRQLDELEGIPVINIQIWFDRKLNSVDGLCFSRSPLLSVYADM      416
          **:.**::** *.******************:::******::** ***************

PDS1      SVTCKEYEDPNKSMLELVFAPCSPIAGGNVNWIGKSDEEIIDATMGELARLFPTEIANDD      473
PDS2      STCCEEYASNDKSMLELVFAPCSPEAGSPLNWIAKPDSDIIDATMKELERLFPLEIGPDA      476
          *. *:** . :************* **. :***.* *.:****** ** **** **. *

PDS1      KWPATKMQGPNGQAKLEKYAVVKVPRSVYAAIPGE-------------------------      508
PDS2      PE--------EKRANVVKSTVVRVPRSVYAAVPGRNKYRPSQESPIENFIMAGDYATQKY      528
                    : :*:: * :**:*********:**.

PDS1      ------------------------------------------------------------      508
PDS2      LGSMEGAVLSGKLAAEVICDKFMGRAERKGVKEVHSSVLTKQIEERTPAGIAMEKGRVSP      588


PDS1      -------------- 508
PDS2      TSYGGGQQGGFENP 602
```

2)β-carotene hydroxylase, BCH

Plant and Green Algal BCH Sequences (non-heme di-iron):

Vitis vinifera, GenBank accession number AAM77007.

```
  1 matgisasln smscrlgrns ftatgpssvi slssfltpvt hlkgnifplq rrrslkvclv
 61 lekeiedgie ieddspessn raserlarkk aerytylvaa mmsslgitsm aivavyyrls
121 wqmeggeipv lemlgtfals vgaavgmefw arwahkalwh aslwhmhesh hrpregpfel
181 ndvfaiinav paisllsygl fnkglvpglc fgaglgitvf gmaymfvhdg lvhrrfpvgp
241 ianvpylrkv asahqlhhsd kfngvpyglf lgpmeleevg gmeelekeis rrikssdss
```

Haematococcus pluvialis (*Haematococcus lacustris*), GenBank accession number AAO53295.

```
  1 ittmlsklqs isvkarrvel arditrpkvc lhaqrcslvr lrvaapqtee avgtqqaaga
 61 gdehsadval qqldraiaer rarrkreqls yqaaaiaasi gvsgiaifat ylrfamhmtv
121 ggavpwgeva gplllvvggq lgmemyarya hkaiwhespl gwllhkshht prtgpfeand
181 lfaiikglpa mllctfgfwl pnvlgtacfg aglgitlygm aymfvhdglv srrfptgpia
241 glpymkrltv ahqlhhsgky ggapwgmflg pqelqhipga aeeverlvle ldwskr
```

Capsicum annuum, CAA70888.

```
  1 ttgryhyqlv wcqisfssts rtsyyrhspf lgpkptpttp svypitpfsp nlgsilrcrr
 61 rpsftvcfvl eddkfktqfe ageediemki eeqisatrla eklarkkser ftylvaavms
121 sfgitsmavm avyyrfywqm eggevpfsem fgtfalsvga avgmefwarw ahkalwhasl
181 whmheshhkp regpfelndv faiinavpai alldygffhk glipglcfga glgitvfgma
241 ymfvhdglvh krfpvgpvan vpylrkvaaa hslhhsekfn gvpyglflgp keleevggle
301 elekevnrrt ryikgs
```

```
T_pseudonana      ------------------------------------------------------------        0
V_vinifera        ---------matgisaslnsmscrlgrnsftatgpssvislssfltpvthlkgnifplqr        51
C_annuum          ttgryhyqlvwcqisfsstsrtsyyrhspflgpkptpttpsvypitpfspnlgsilrcrr        60
H_pluvialis       -----------------------------------ittmlsklqsisvkarrvelard        23


T_pseudonana      ----------------------------------------------------MHQLSR        6
V_vinifera        rrslkvclvle--------------kei--------ed-gieieddspessnraserlar        88
C_annuum          rpsftvcfvle--------------ddkfktqfeagee-diemkieeqisatrlaeklar        105
H_pluvialis       itrpkvclhaqrcslvrlrvaapqteeavgtqqaagagdehsadvalqqldraiaerrar        83
                                                                      .: :*


T_pseudonana      LLSKLFPVIATYIIAKYALPHINTI---LH---------CSNTLSNFVRITYTVLFAIAM        54
V_vinifera        kkaerytylvaammsslgitsmaivavyyrlswqme-ggeipvlemlgtfalsvgaavgm        147
C_annuum          kkserftylvaavmssfgitsmavmavyyrfywqme-ggevpfsemfgtfalsvgaavgm        164
H_pluvialis       rkreqlsyqaaaiaasigvsgiaifatylrfamhmtvggavpwgevagplllvvggqlgm        143
                     :     .: : :. .:  :  .    :              .    :   *   :.*


T_pseudonana      EYISRYSHCYLWHGKFLWWINGSHHHQYPAVGSTPVYGHNNPYVSPAIELNDAFAVFFAT        114
V_vinifera        efwarwahkalwhas--lwhmheshhr-pr--------------egpfelndvfaiinav        190
C_annuum          efwarwahkalwhas--lwhmheshhk-pr--------------egpfelndvfaiinav        207
H_pluvialis       emyaryahkaiwhesplgwllhkshht-pr--------------tgpfeandlfaiikgl        188
                  *  :*::*  :** .   *   . **  *                 :* ** **:: .


T_pseudonana      IATLAMWIGSEPPSTLTKDCSIGIGLGVTLYGLSYFVGHDIVAHERLGKGVANALRRAFP        174
V_vinifera        paislsyglfnkglv-pglcfgaglgitvfgmaymfvhdglvhrrfpvgpia----nvp        245
C_annuum          paialldygffhkgli-pglcfgaglgitvfgmaymfvhdglvhkrfpvgpva----nvp        262
H_pluvialis       pamllctfgfwlpnvl-gtacfgaglgitlygmaymfvhdglvsrrfptgpia----glp        243
                   *      *    . :    .:* ***:*::*::*:. ** :. .*:  *        .*


T_pseudonana      YMEQCASVHIRYHHKLTKRSNDSDPYGAPYGFWLGPSEVECLNRGQWYAPMPMSLKAISW        234
V_vinifera        ylrkvasahqlhh--------sdkfngvpyglflgpmeleevggmeele--keisrriks        295
C_annuum          ylrkvaaahslhh--------sekfngvpyglflgpkeleevggleele--kevnrrtry        312
H_pluvialis       ymkrltvahqlhh--------sgkyggapwgmflgpqelqhipgaaeev--erlvleldw        293
                  *:.: : .*  :*        . .  *.*:*::*** *:: :


T_pseudonana      IATLIFFASTIHSSLSPAAQAIVLLGCVGWCGSGTLSDNTTTSRRIGKLLSLNWQSTPSR        294
V_vinifera        sdss----------------------------------------------------------        299
C_annuum          ikgs----------------------------------------------------------        316
H_pluvialis       skr-----------------------------------------------------------        296


T_pseudonana      LMPHGLSGLISVGIGSYLIFGHSLVGDLKPYTMQQPPYLIILYATATSWNALGGYMIVNT        354
V_vinifera        ----------------------------------------------------------          299
C_annuum          ----------------------------------------------------------          316
H_pluvialis       ----------------------------------------------------------          296


T_pseudonana      APPNTRMLFRRCAILQVCLSYFIVRFLPHSSVLLIRLESNTITTSLRCLDLIVTISAVVC        414
V_vinifera        ----------------------------------------------------------          299
C_annuum          ----------------------------------------------------------          316
H_pluvialis       ----------------------------------------------------------          296


T_pseudonana      TLSFFDAVVDMSKQSVVLGQSIAFGIIGILLLSVYPIQLSLQGEEWWSCIQNRYPMQASG        474
V_vinifera        ----------------------------------------------------------          299
C_annuum          ----------------------------------------------------------          316
H_pluvialis       ----------------------------------------------------------          296
```

```
T_pseudonana      MIAYIYVPATVTFSLFLFGATLYQRKIMSASEYGIISLMVILVCLLATVLSQEIHIPDVS        534
V_vinifera        ------------------------------------------------------------        299
C_annuum          ------------------------------------------------------------        316
H_pluvialis       ------------------------------------------------------------        296

T_pseudonana      TQRIYLPCEDPAMDSLEEKVLEALDFSRYARSILTTVLGIKFESPA   580
V_vinifera        ---------------------------------------------   299
C_annuum          ---------------------------------------------   316
H_pluvialis       ---------------------------------------------   296
```

T. pseudonana BCH **(Appendix C)**

MHQLSRLLSKLFPVIATYIIAKYALPHINT
ILHCSNTLSNFVRITYTVLFAIAMEYISRYSHCYLWHGKFLWWINGSHHHQYPAVGSTPV
YGHNNPYVSPAIELNDAFAVFFATIATLAMWIGSEPPSTLTKDCSIGIGLGVTLYGLSYF
VGHDIVAHERLGKGVANALRRAFPYMEQCASVHIRYHHKLTKRSNDSDPYGAPYGFWLGP
SEVECLNRGQWYAPMPMSLKAISWIATL IFFASTIHSSLSPAAQAIVLLGCVGWCGSGTL
SDNTTTSRRIGKLLSLNWQSTPSRLMPHGLSGLISVGIGSYLIFGHSLVGDLKPYTMQQP
PYLIILYATATSWNALGGYMIVNTAPPNTRMLFRRCAILQVCLSYFIVRFLPHSSVLLIR
LESNTITTSLRCLDLIVTISAVVCTLSFFDAVVDMSKQSVVLGQSIAFGIIGILLLSVYP
IQLSLQGEEWWSCIQNRYPMQASGMIAYIYVPATVTFSLFLFGATLYQRKIMSASEYGII
SLMVILVCLLATVLSQEIHIPDVSTQRIYLPCEDPAMDSLEEKVLEALDFSRYARSILTT
VLGIKFESPA-

The underlined portion is what corresponds to BCHs from other organisms in the above alignment.

Percent Identity Matrix

```
1: T_pseudonana  100.00   25.73   26.21   27.05
2: V_vinifera     25.73  100.00   68.56   42.96
3: C_annuum       26.21   68.56  100.00   42.45
4: H_pluvialis    27.05   42.96   42.45  100.00
```

Cyanobacterial BCH Sequences (non-heme di-iron):

Prochlorococcus marinus subsp. marinus str. CCMP1375, GenBank accession number AAP99312.

```
  1 mtntlnttih ndlppkfqss sfwqkrikdy ldppnffnpt lglfiggyai aflsiwqwyk
 61 gvwplpvlvg laflslhmeg tvihdachka ahpnkwinqa mghgaaillg fsfpvftrvh
121 lqhhshvndp kndpdhivst fgpvwliapr ffyheyfffq rklwrkyelm qwglersifi
181 tivlagvhfn fmnviynlwf gpalmvgvtl giffdylphr pfmarnkwkn srvypsrvmn
241 ilimgqnyhl vhhlwpsipw feykpayeat kplldqkgsp qrmgifeskk dsfnflydii
301 lgirshkksr skmrplanli ptkklrrkwl yilhktaiip dkid
```

Thermosynechococcus elongatus BP-1, GenBank accession number NP_682690.

```
  1 miteaaipat vpkeflgppv gfnptlvmff aafaiailst wgyvqghwpg glsfvanmla
 61 lhlmgtvihd ashnvahrhp imnaimghgs almlgfvfpv ftrvhmqhha hvndpendpd
121 hyvskggplw liaprffyhe ifffkrrlwr kyellewfls rltlvgivtf aalngyldyi
181 lnywflpagv vglmlglffd yfphrphter drwhnarvyp srllnilifg qnyhlihhlw
241 pnipwykvqp ayyavkplld ahgckqtlgi lepgnflpfl ydalvglhfh rprss
```

Cyanidioschyzon merolae strain 10D, GenBank accession number NP_848964.

```
  1 mnsllfflsv sltlvsligy vmhmplsvcf vfnvislhla gslihdashk sahsneying
 61 iighvcgfll gfsfvvfkkv hmqhhahvnq akydpdhyvs tggpiwliap rffyheiyff
121 qrrlyrnhel lewmmarglf flvlmaawkf galtyvlrcw fcaallvgtf lglcfdylph
181 ypfvqthrwh naciqendml nwlilgqnyh lvhhlwpsep wykyqqkyqa hqqlftpqtc
241 vlgwptqigy dlcfglrig
```

```
T_pseudonana      MHQLSRLLSKLFPVIATYIIAKYALPHINTILHCSNTLSNFVRITYTVLFAIAMEYISRY        60
C_merolae         ------------------------------------------------------------        0
P_marinus         ------------------------------------------------------------        0
T_elongatus       ------------------------------------------------------------        0


T_pseudonana      SHCYLWHGKFLWWINGSHHHQYPA-VGSTPV-YGHNNPYVSPAIELNDAFAVFFATIAT-        117
C_merolae         ----------------------------------------------mnsllfflsvslt        13
P_marinus         ---------mtntlnttihndlppkfqsssfwqkrikdyldppnffnptlglfiggyaia        51
T_elongatus       ------------------------miteaaipatvpkeflgppvgfnptlvmffaafaia        36
                                                                  :*:   :


T_pseudonana      -LAMW--I-GSEPPSTLTKDCSIGIGLGVTLYGLSYFVGH------DIVAHERLGKGVAN        167
C_merolae         lvsligyv-mhmplsvcfvfnv----islhlagslihdashksahsneyingiighvcgf        68
P_marinus         flsiwqwykgvwplpvlvglaf----lslhmegtvihdachkaahpnkwinqamghgaai        107
T_elongatus       ilstwgyvqghwpgglsfvanm----lalhlmgtvihdashnvahrhpimnaimghgsal        92
                   ::         *             :.: : *   . .       .   :  :*:  .


T_pseudonana      ALRRAFPYMEQCASVHIRYHHKLTKRSNDSDPYGAPYG-FWLGPS-----EVECLNRGQW        221
C_merolae         llgfsfvvf---kkvhmqhhahvnqakydpdhyvstggpiwliaprffyheiyffqrrly        125
P_marinus         llgfsfpvf---trvhlqhhshvndpkndpdhivstfgpvwliaprffyheyfffqrklw        164
T_elongatus       mlgfvfpvf---trvhmqhhahvndpendpdhyvskggplwliaprffyheifffkrrlw        149
                   *   *  :     **:::* ::.. . * *   :  * .**        *   ::*  :


T_pseudonana      YAPMPMSLKAISWIATLIFFASTIHSSLSPAAQAIVLLGCVGWCGSGTLSDNTTTSRRIG        281
C_merolae         rn-----hellewmmarglfflvlmaawkfgalt-yvl--rcwfcaal---------lvg        168
P_marinus         rk-----yelmqwglersifitivlagvhfnfmn-viy--nlwfgpal---------mvg        207
T_elongatus       rk-----yellewflsrltlvgivtfaalngyld-yil--nywflpag---------vvg        192
                         : :.*      :   :  .         :     *   .           :*


T_pseudonana      KLLSLNWQSTP------------SRLMPHGLSGLISVGIGSYLIFGHSL--VGDL---KP        324
C_merolae         tflglcfdylphypfvqthrwhnaciqendmln--------wlilgqnyhlvhhlwpsep        220
P_marinus         vtlgiffdylphrpfmarnkwknsrvypsrvmn--------ilimgqnyhlvhhlwpsip        259
T_elongatus       lmlglffdyfphrphterdrwhnarvypsrlln--------ilifgqnyhlihhlwpnip        244
                    *.: ::  *            : :    : .         **:*:.   : .*    *


T_pseudonana      YTMQQPPYLIILYATATSWNALGG---YMIVNTAPPNTRMLFRRCAILQVCLSYFIVRFL        381
C_merolae         wykyqqkyqah----qqlftpqtcvlgwpt--------qigydlc------fgl---rig        259
P_marinus         wfeykpayeat----kplldqkgspqrmgifeskkdsfnflydii------lgi---rsh        306
T_elongatus       wykvqpayyav----kplldahgckqtlgilepg-nflpflydal------vgl---hfh        290
                  :   :  *                                 : :         ..    :


T_pseudonana      PHSSVLLIRLESNTIT-TSLRCLDLIVTISAVVCTLSFFDAVVDMSKQSVVLGQSIAFGI        440
C_merolae         ------------------------------------------------------------        259
P_marinus         kksrskmrpl-anliptkklrrkwlyi------------------lhktaiipdkid---        344
T_elongatus       rprss--------------------------------------------------------        295


T_pseudonana      IGILLLSVYPIQLSLQGEEWWSCIQNRYPMQASGMIAYIYVPATVTFSLFLFGATLYQRK        500
C_merolae         ------------------------------------------------------------        259
P_marinus         ------------------------------------------------------------        344
T_elongatus       ------------------------------------------------------------        295


T_pseudonana      IMSASEYGIISLMVILVCLLATVLSQEIHIPDVSTQRIYLPCEDPAMDSLEEKVLEALDF        560
C_merolae         ------------------------------------------------------------        259
```

```
P_marinus       ------------------------------------------------------------       344
T_elongatus     ------------------------------------------------------------       295


T_pseudonana    SRYARSILTTVLGIKFESPA580
C_merolae       --------------------259
P_marinus       --------------------344
T_elongatus     --------------------295


Percent Identity Matrix:

1: T_pseudonana  100.00   17.49   21.71   22.27
2: C_merolae      17.49  100.00   45.17   49.81
3: P_marinus      21.71   45.17  100.00   53.90
4: T_elongatus    22.27   49.81   53.90  100.00
```

Non-Photosynthetic Eubacterial BCH Sequences (non-heme di-iron):

Pantoea stewartia, GenBank accession number AAN85601.

```
  1 mlwiwnaliv fvtvvgmevv aalahkyimh gwgwgwhlsh heprkgafev ndlyavvfai
 61 vsialiyfgs tgiwplqwig agmtayglly fmvhdglvhq rwpfryiprk gylkrlymah
121 rmhhavrgke gcvsfgflya pplsklqatl rerhaarsga ardeqdgvdt ssgk
```

Pseudomonas putida, GenBank accession number KT2440 NP_745389.

```
  1 mlfnlailfg tlvamegvgt lahkyimhgw gwwlhrshhe phlgmletnd lylvalglia
 61 talvalgksg yaplqwvggg vagygalyvl ahdgffhrhw prkprpvnry lkrlhrahrl
121 hhavkgrtgs vsfgffyapp lkvlkqqlrs rrsqs
```

Flavobacterium sp. ATCC 21588, GenBank accession number AAC44852.

```
  1 mstwaailtv iltvaamelt aysvhrwimh gplgwgwhks hhdedhdhal ekndlygvif
 61 avisivlfai gamgsdlaww lavgvtcygl iyyflhdglv hgrwpfryvp krgylrrvyq
121 ahrmhhavhg rencvsfgfi wapsvdslka elkrsgallk dregadrnt
```

```
T_pseudonana        MHQLSRLLSKLFPVIATYIIAKYALPHINTILHCSNTLSNFVRITYTVLFAIAMEYISRY      60
Flavobacterium      -------------------------------------mstwaailtviltvaameltays      23
P_stewartia         -------------------------------------mlwiw-nalivfvtvvgmevvaal      23
P_putida            -------------------------------------mlf-nlailfgtlvamegvgtl      21
                                                          :     .    .**  .

T_pseudonana        SHCYLWHGKFLWWINGSHHHQYPAVGSTPVYGHNNPYVSPAIELNDAFAVFFATIATLAM      120
Flavobacterium      vhrwimhgplgwgwhkshhded---------------hdhalekndlygvifavisivlf      68
P_stewartia         ahkyimhg-wgwgwhlshhe-p---------------rkgafevndlyavvfaivsiali      66
P_putida            ahkyimhg-wgwwlhrshhe-p---------------hlgmletndlylvalgliatalv      64
                     * :: **   *  : ***.                   :* ** : * :. ::   .

T_pseudonana        WIGSEPPSTLTKDCSIGIGLGVTLYGLSYFVGHDIVAHERLGKGVANALRRAFPYMEQCA      180
Flavobacterium      aigamgsd-----lawwlavgvtcygliyyflhdglvhgrwpfryvpk----rgylrrvy      119
P_stewartia         yfgstgiw-----plqwigagmtayglllyfmvhdglvhqrwpfryipr----kgylkrly      117
P_putida            algksgya-----plqwvgggvagygalyvlahdgffhrhwprkprpv----nrylkrlh      115
                     :*             :. *:: **  * . ** . * :             *:.:

T_pseudonana        SVHIRYHHKLTKRSNDSDPYGAPYGFWLGPSEVECLNRGQWYAPMPMSLKAISWIATLIF      240
Flavobacterium      q-ahrmhhavhgr-------------------encvsfgfiwapsvdslkaelk------      153
P_stewartia         m-ahrmhhavrgk-------------------egcvsfgflyapplsklqatlr------      151
P_putida            r-ahrlhhavkgr-------------------tgsvsfgffyapplkvlkqqlr------      149
                        * ** :  :                    .:. *  :**    *:
```

```
T_pseudonana        FASTIHSSLSPAAQAIVLLGCVGWCGSGTLSDNTT--TSRRIGKLLSLNWQSTPSRLMPH      298
Flavobacterium      ------rsga------------------llkdregadrnt------------------      169
P_stewartia         ------erhaa--------------rsgaardeqdgvdtsssgk----------------      175
P_putida            ------srrsq--------------s----------------------------------      155
                           :

T_pseudonana        GLSGLISVGIGSYLIFGHSLVGDLKPYTMQQPPYLIILYATATSWNALGGYMIVNTAPPN      358
Flavobacterium      ------------------------------------------------------------      169
P_stewartia         ------------------------------------------------------------      175
P_putida            ------------------------------------------------------------      155


T_pseudonana        TRMLFRRCAILQVCLSYFIVRFLPHSSVLLIRLESNTITTSLRCLDLIVTISAVVCTLSF      418
Flavobacterium      ------------------------------------------------------------      169
P_stewartia         ------------------------------------------------------------      175
P_putida            ------------------------------------------------------------      155


T_pseudonana        FDAVVDMSKQSVVLGQSIAFGIIGILLLSVYPIQLSLQGEEWWSCIQNRYPMQASGMIAY      478
Flavobacterium      ------------------------------------------------------------      169
P_stewartia         ------------------------------------------------------------      175
P_putida            ------------------------------------------------------------      155


T_pseudonana        IYVPATVTFSLFLFGATLYQRKIMSASEYGIISLMVILVCLLATVLSQEIHIPDVSTQRI      538
Flavobacterium      ------------------------------------------------------------      169
P_stewartia         ------------------------------------------------------------      175
P_putida            ------------------------------------------------------------      155


T_pseudonana        YLPCEDPAMDSLEEKVLEALDFSRYARSILTTVLGIKFESPA     580
Flavobacterium      ---------------------------------------     169
P_stewartia         ---------------------------------------     175
P_putida            ---------------------------------------     155

1: T_pseudonana    100.00   29.34   28.90   27.74
2: Flavobacterium   29.34  100.00   50.60   41.18
3: P_stewartia      28.90   50.60  100.00   52.90
4: P_putida         27.74   41.18   52.90  100.00
```

Archaeal BCH Sequence (non-heme di-iron):

Sulfolobus solfataricus P2, GenBank accession number NP_344225.

```
  1 mmliyyvgma vltfvgmefv arlmhkyvmh gllwfihedh hkekqaelek ndlfglvfas
 61 vsvylfflgi qgsyvalsia igmssygiay ffihdmvihd rhlhlrswgl khrpfkdlil
121 vhdihhkegk gnwgflfvik gldkvpilkd e

S_solfataricus      ------------------------------------mmliyyvgmavltfvgmefvarl      23
T_pseudonana        MHQLSRLLSKLFPVIATYIIAKYALPHINTILHCSNTLSNFVRITYTVLFAIAMEYISRY      60
                                                       :  :  :  :**  :.**:::*

S_solfataricus      mhkyvmhg-llwfihedhhkekq----------------aeleknd lfglvfasvsvylf      66
T_pseudonana        SHCYLWHGKFLWWINGSHHHQYPAVGSTPVYGHNNPYVSPAIELNDAFAVFFATIATLAM      120
                     * *: ** :**:*: .**::                   :* ** *.:.**:::.  :

S_solfataricus      flgiqgs-----yvalsiaigmssygiayffihdmvihdrhlhlrswglkhrpfkdlilv      121
T_pseudonana        WIGSEPPSTLTKDCSIGIGLGVTLYGLSYFVGHDIVAHERLGKGVANALR-RAFPYMEQC      179
                    ::* :         ::.*.:*:: **::**. **:* *:*  :  : .*: * *  :

S_solfataricus      hd----ihhke---------gkgnwgflfvikgldkvpilkde-----------------      151
T_pseudonana        ASVHIRYHHKLTKRSNDSDPYGAPYGFWLGP---SEVECLNRGQWYAPMPMSLKAISWIA      236
                     .     ***           . :** :     .:*  *:
```

```
S_solfataricus      ------------------------------------------------------------      151
T_pseudonana        TLIFFASTIHSSLSPAAQAIVLLGCVGWCGSGTLSDNTTTSRRIGKLLSLNWQSTPSRLM      296

S_solfataricus      ------------------------------------------------------------      151
T_pseudonana        PHGLSGLISVGIGSYLIFGHSLVGDLKPYTMQQPPYLIILYATATSWNALGGYMIVNTAP      356

S_solfataricus      ------------------------------------------------------------      151
T_pseudonana        PNTRMLFRRCAILQVCLSYFIVRFLPHSSVLLIRLESNTITTSLRCLDLIVTISAVVCTL      416

S_solfataricus      ------------------------------------------------------------      151
T_pseudonana        SFFDAVVDMSKQSVVLGQSIAFGIIGILLLSVYPIQLSLQGEEWWSCIQNRYPMQASGMI      476

S_solfataricus      ------------------------------------------------------------      151
T_pseudonana        AYIYVPATVTFSLFLFGATLYQRKIMSASEYGIISLMVILVCLLATVLSQEIHIPDVSTQ      536

S_solfataricus      -------------------------------------------   151
T_pseudonana        RIYLPCEDPAMDSLEEKVLEALDFSRYARSILTTVLGIKFESPA   580
```

Percent Identity Matrix:

```
1: S_solfataricus  100.00   28.57
2: T_pseudonana     28.57  100.00
```

Cytochrome 450-type BCH Sequences:

Thermus thermophilus HB8, GenBank accession number BAD71899.

```
  1 mkrlslreaw pylkdlqqdp lavllewgra hprlflplpr fplalifdpe gvegallaeg
 61 ttkatfqyra lsrltgrgll tdwgkswkea rkalkdpflp ksvrgyream eeeawaffge
121 wrgeerdldh emlalslrll gralfgkpls pslaehalka ldrimaqtrs plalldlaae
181 arfrkdrgal yreaealivh pplshlprer alseavtllv aghetvasal twsflllshr
241 pdwqkrvaes eeaalaafqe alrlyppawi ltrrlerpll lgedrlpqgt tlvlspyvtq
301 rlyfpegeaf qperflaerg tpsgryfpfg lgqrlclgrd fallegpivl raffrrfrld
361 plpfprvlaq vtlrpegglp arpregvra
```

Xanthophyllomyces dendrorhous, GenBank accession number CDN65464.

```
  1 mfilvlltga lglaafswas iaffslylap rrsslynlqg pnhtnyftgn fldilsartg
 61 eehakyreky gstlrfagia gapvlnstdp kvfnhvmkea ydypkpgmaa rvlriatgdg
121 vvtaegeahk rhrrimipsl saqavksmvp iflekgmelv dkmmedaaek dmavgesage
181 kkatrleteg vdvkdwvgra tldvmalagf dyksdslqnk tnelyvafvg ltdgfaptld
241 sfkaimwdfv pyfrtmkrrh eipltqglav srrvgielme qkkqavlgsa sdqavdkkdv
301 qgrdilsllv raniaanlpe sqklsdeevl aqisnllfag yetsstvltw mfhrlsedka
361 vqdklreeic qidtdmptld elnalpylea fvkeslrldp pspyanrecl kdedfiplae
421 pvigrdgsvi nevritkgtm vmlplfninr skfiygedae efrperwled vtdslnsiea
481 pyghqasfis gpracfgwrf avaemkaflf vtlrrvqfep iishpeyehi tliisrpriv
541 grekegyqmr lqvkpve
```

```
T_pseudonana        MHQLSRLLSKLFPVIATYIIAKYALPHINTILHCSNTLSNF--VRITYTVLFAI----AM      54
T_thermophilus      ------------------------------------------------------------      0
X_dendrorhous,      ----------mf------il---------vlltgalglaafswasiaffslylaprrssl      35
```

```
T_pseudonana        EYISRYSHCYLWHGKFLWWIN---GSHHHQY----------PAVGSTPVYGHNNPYVSPA      101
T_thermophilus      ----------------------------------------------------mkrl      4
X_dendrorhous,      ynlqgpnhtnyftgnfldilsartgeehakyrekygstlrfagiagapvlnstdpkvfnh      95
                                                                           :

T_pseudonana        IELNDAFAVFFATIATLAMWIGSEPPSTLTKDCSIGIGLGVTLYGLSYFVGHDIVAHERL      161
T_thermophilus      -slreawpy-------------lk-dlqqdp-------lavlle-----wgr---ahprl      34
X_dendrorhous,      -vmkeaydy-------------pk-pgmaarvlriatgdgvvta-----egeahkrhrri      135
                      :.:*:               :              .*        *.    * *:

T_pseudonana        GKGVA--NALRRAF------------PYMEQCASVHIRYHHKL-TKRSNDSDPYGAPYGF      206
T_thermophilus      flplprf----plalifdpegveg--allaegttka-----tfqyralsrltgrg-lltd      82
X_dendrorhous,      mipslsaqavksmvpiflekgmelvdkmmedaaekdmavgesagekkatrletegvdvkd      195
                                         : : :       .   :  .     *

T_pseudonana        WLGPSEVECLNRGQ--WY----------------------AP-----MPMSLKAISWIAT      237
T_thermophilus      wgks------------wkea------------rkalkdpflpksvrgyreameeeawaff      118
X_dendrorhous,      wvgratldvmalagfdyksdslqnktnelyvafvgltdgfap-----tldsfkaimwdfv      250
                    *              :                          *       :::   *

T_pseudonana        LIFFASTIHSSLSPAAQAIVLLGCVGWCGSGTLSDNTTTSRRIGKLLSLNWQSTPSRLMP      297
T_thermophilus      gewrgeerd-ldhem---l------------------alslrllgral-----------fg      146
X_dendrorhous,      pyfrtmkrr-heipltqgl------------------avs-rrvgiel-----------me      280
                      :              :                 : : * :*  *           :

T_pseudonana        HGLSGLISVGIGSYLIFGHSLVGDLKPYTMQQPPYLI-ILYATATSWNALGGYMIV----      352
T_thermophilus      kpls----------pslaehalkaldrimaqtrsplalldlaaearfrkdrgalyreaea      196
X_dendrorhous,      qkkq----------avlgsasdqavdkkdvqgrdils----------------llvrani      314
                    :  .           :.      :.    *    *                 :

T_pseudonana        -NTAPPNTRM------------LFRRCAILQVCLSYFIVRFLPHSSVLLIRL--ESNTIT      397
T_thermophilus      livhpplshlpreralseavtllvaghetvasaltwsflll-shrpdwqkrvaeseeaa-      254
X_dendrorhous,      aanlpesqklsdeevlaqisnllfagyetsstvltwmfhrl-sedkavqdklreeicqid      373
                        *   ::            *.         *:: :  :  .      ::  .

T_pseudonana        TSLRCLDLIVTISAVVCTLSFFDAVVDMSKQ-------------------------SVVL      432
T_thermophilus      ------------------laafqealrlyppawiltrrl---------er------plll      281
X_dendrorhous,      tdmptldel---nalpyleafvkeslrldppspyanreclkdedfiplaepvigrdgsvi      430
                                    : ..  : :                              ::

T_pseudonana        G-QSIAFGIIGILLLSVYPIQ----------LSLQGEEWWSCIQNRYPMQASGMIAYIYV      481
T_thermophilus      gedrlpqg--ttlvlspyvtq---rlyfpegeafqperflaer---------------gt      321
X_dendrorhous,      nevritkg--tmvmlplfninrskfiygedaeefrperwledvtdsl--------nsiea      480
                    .   :  *    ::*  :  :           :: *.:                    .

T_pseudonana        PATVTFSLFLFGATLYQRKIMSASEYGIISLMVILVCLLATVLSQEIHIPDVSTQRIYLP      541
T_thermophilus      psgryfp-fglgqrlcl-----grdfallegpivlraffrrfrldplpfprvla-qvtlr      374
X_dendrorhous,      pyghqas-fisgpracf-----gwrfavaemkaflfvtlrrvqfepiishpeye-hitli      533
                    *       *  *          .  :.: .   .*   :  .  : :        :: *

T_pseudonana        CEDPAMDSLE-EKVLEALDFSRYARSILTTVLGIKFESPA       580
T_thermophilus      -pegglparpregvra-----------------------        389
X_dendrorhous,      isrprivgrekegyqmrlqvkpve----------------       557
                         : .   *
```

Percent Identity Matrix:

```
1: T_pseudonana    100.00   16.77   20.09
2: T_thermophilus   16.77  100.00   26.98
3: X_dendrorhous_   20.09   26.98  100.00
```

3)LUT1-like, LTL

Peptide alignment of the two *T. pseudonana* LTLs.

```
LTL1      --MKFTT--ALA--VLCWTSVTNAFVPSSFTSPA------LKNEQQQVRASSPLYALDTK      48
LTL2      MCTKLSSRRTLLALYFAFTGCTAFQLPSATPSRASITKAYSTHLDKEIKSKTPLVNPSKI      60
             *:::  :*    :.:*. *   :**:  * *       .: :::::::.:**   ..

LTL1      EKEETTTATSASSTDTSSTPAA--AATEESEGLPWWWEYI--------------WK----      88
LTL2      YTQADIDTLDLSSYEN-ELLAAWDTDSSLQRGFDWEIEKLRRNFAGLRQREDGQWVRKPS      119
           .:    : . ** :. .  **  : :. ..*: *  * :              *

LTL1      -LPVMQPAE----------------PGTDIIFADSARVLRT-------------------      112
LTL2      LFDFLVTNTPSNVVGVSNTGERYESPPKPVNMLDVGLLITKNLLNTLGFGPSLGMAAVPD      179
           : .:                    * . : : * . :: .

LTL1      -NIEQIYGGFPSLDQCPLAEGEITDIADGTMFIGLQRYQQQYGSPYKLCFGPKSFLVISD      171
LTL2      AVIQKYEGSFFSF-IKGVLGGDLQTLAGGPLFLLLAKYYQDYGPIFNLSFGPKSFLVISD      238
            *::  *.* *:    :  *::  :*.* :*: * :* *:**  ::*.***********

LTL1      PVQAKHVLRDAN-TLYDKGILAEILKPIMGKGLIPADPETWSVRRRAIVPAFHKAWLNHM      230
LTL2      PVMARHILRDSSPEQYCKGMLAEILEPIMGDGLIPADPKIWKVRRRAVVPGFHKKWLNNM      298
          ** *:*:***:.   * **:*****:****.*******: *.*****:**.*** ***:*

LTL1      VGLFGYCNEGLIASLEEAAKKNDAPNGQQGGKIEMEEKFCSVALDIIGLSVFNYEFGSVS      290
LTL2      VTLFGDCGERLVNDLDARAT--------AKTPVDMEERFCSVTLDIIGKAVFNYDFGSVT      350
          * *** *.* *: .*:  *.           ::***:****:***** :****:****:

LTL1      EESPVIKAVYSALVEAEHRSMTPAPYWDLPFANEVVPRLRKFNSDLKVLDDVLTDLIDRA      350
LTL2      KESPIVKAVYRVLREAEHRSSSFIPYWDLPYADKWMGGQVEFRKDMGMLDDILTKLINRA      410
          :***::**** .* ****** :  ******:*:: :    :*..*: :***:**.**:**

LTL1      KNSRQVEDIEELEKRDYANVKDPSLLRFLVDMRGADIDNKQLRDDLMTMLIAGHETTAAV      410
LTL2      IETRDEASVEELEDRDVG--DDPSLLRFLADMRGEDLTSKVLRDDLMTMLIAGHETTAAM      468
           ::*:  .:****.** .  .********.**** *: .* ******************:

LTL1      LTWALFELTKHPE-QMAKVRAEIDSVLGDR-TPTYDDIKEMQYLRLVVAETLRLYPEPPL      468
LTL2      LTWTVFGLVSNDSGLMKEIQAEVRTVMGDKLRPDYDDIAKMKKMRYALIEALRLYPEPPV      528
          ***::* *..: .  * :::**: :*:**:  * **** :*: :* .: *:********:

LTL1      LIRRCRTENKLPKGGGR---EATVIRGMDIFLSLYNLHHDERFWPEPNEFKPERWESKYI      525
LTL2      LIRRARSEDNLPAGGSGLSGGVKVLRGTDIFISTWNLHRAPEYWENPEKYDPTRWERRFK      588
          ****.*:*::** **.     ..*:** ***:* :***:  .:* :*::::.* *** ::

LTL1      NPEVPEWAGYDPAKWINTNLYPNEVASDFAYLPFGGGARKCVGDEFATLEATVTLAMLLR      585
LTL2      NPGVKGWNGYDPEKQSESSLYPNEITADYAFLPFGAGKRKCIGDQFAMLEASVTLAMIIN      648
          ** *  * **** *  ::.*****:::*:*:****.* ***:**:** ***:*****::.

LTL1      RFEFEFDSAKLAASKIDIMDHPEDLEHAVGMRTGATIHTRKGLHMVIRKREL--------      637
LTL2      KFDFTLVGS----------------PKDVGMKTGATIHTMNGLNLVVSRRSEDNPIPETN      692
          :*:* : .:                : ***:******* :**::*: :*.

LTL1      ------------------------------------------      637
LTL2      DYWIQQHLSRGLNVNGRPYSTNEDAAWTASSRDKNEGVVSRLVN      736


Percent Identity Matrix

1: LTL1        100.00   47.78
2: LTL2         47.78  100.00
```

APPENDIX 2.E PLASMID MAPS AND PRIMERS

1) VDL2 overexpression, nitrate reductase promoter and terminator.

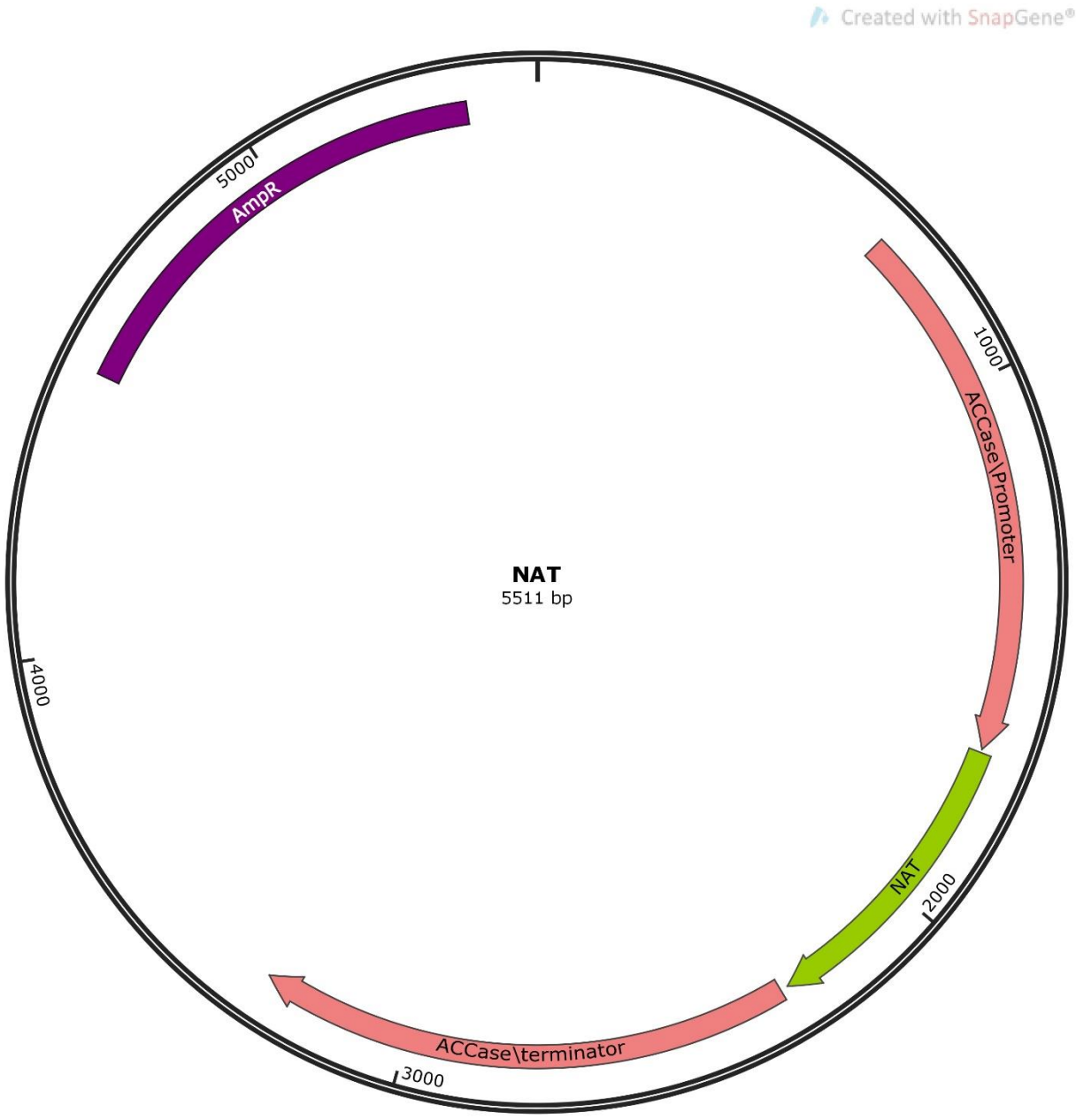
Primers for amplifying the VDL2 open reading frame:

Fwd: 5'ggggacaagtttgtacaaaaaagcaggctATGTCGGCATCATCATCATCAAC3'

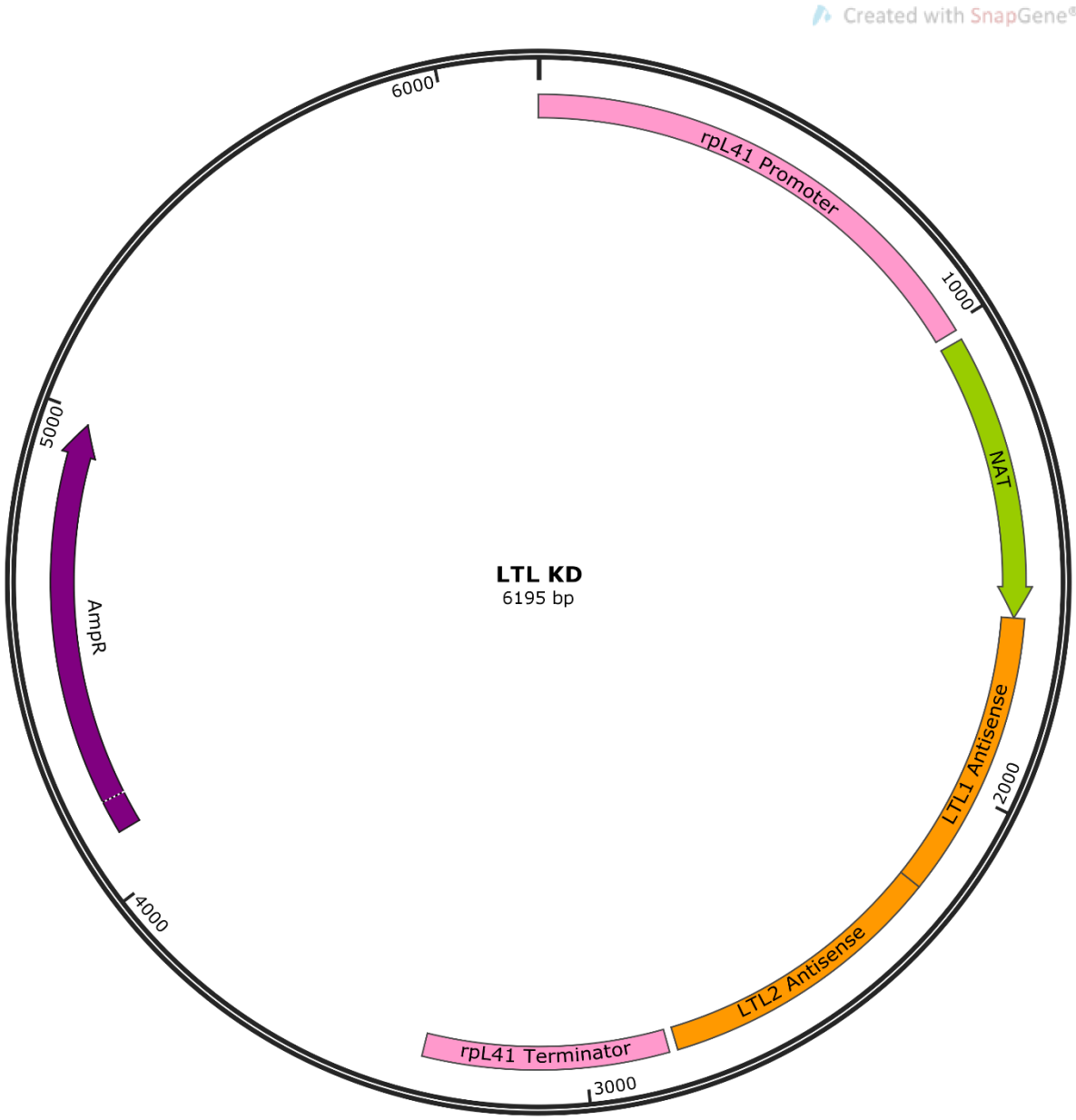
Rev: 5'ggggaccactttgtacaagaaagctgggtaTCACCAATTCTTCTCTTTGATATAAT3'

Lower case sequences add Att sites for Gateway cloning.

2) NAT, acetyl CoA carboxylase promoter and terminator.



3) LTL KD, ribosomal protein 41 promoter and terminator, NAT on the same transcript.



Primers to amplify antisense regions:

Fwd, LTL1: 5'ggggacaagtttgtacaaaaaagcaggctTCCAATGACCATAGTTGTTG3'

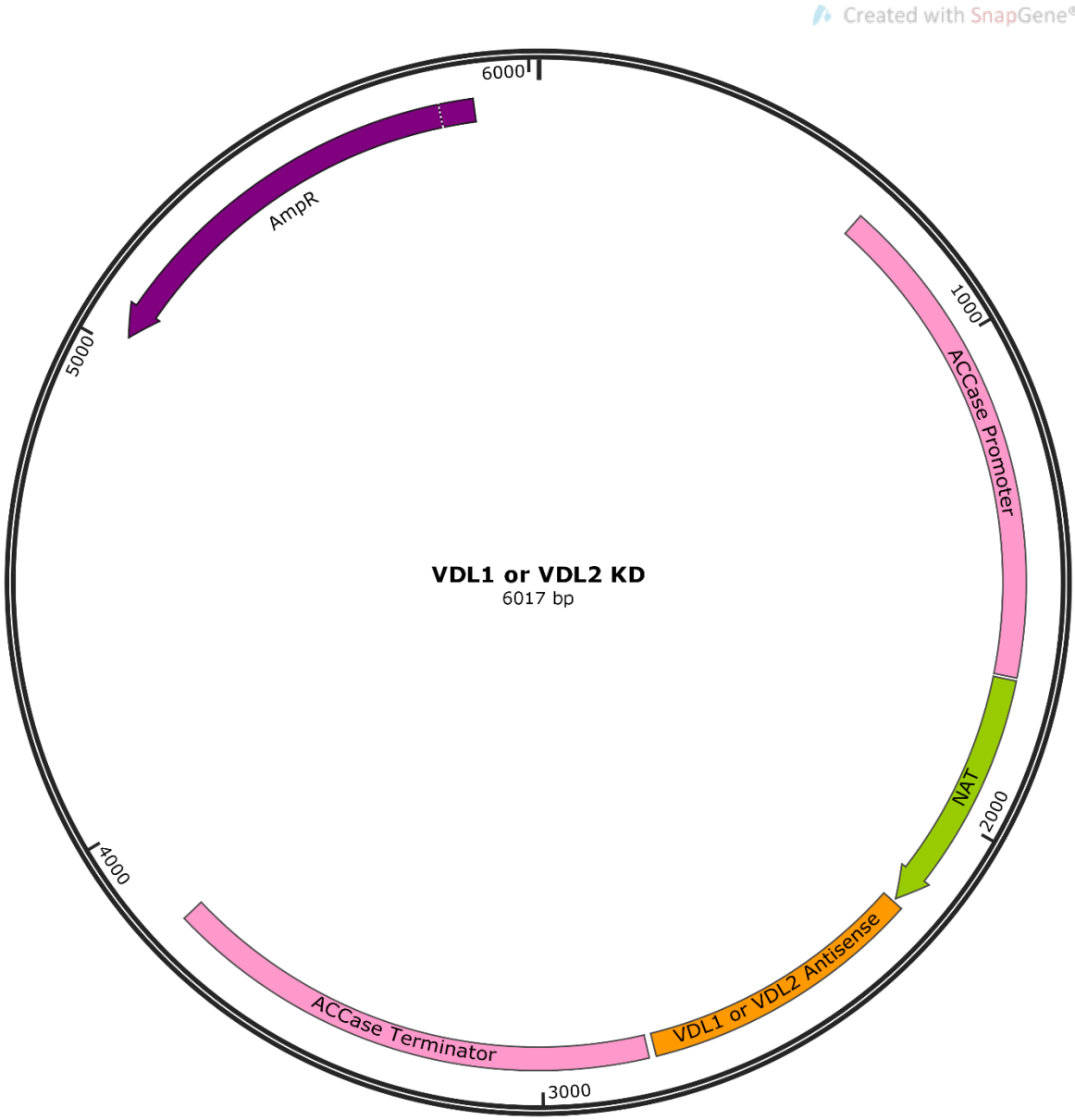
Rev, LTL1: 5'ggggacaacttttgtatacaaagttgTGCAGAGACTTTGAGGTTG3'

Fwd, LTL2: 5'ggggacaactttgtatacaaaagttgCTCTTATATCCTAGATAACTTT3'

Rev, LTL2: 5'ggggaccactttgtacaagaaagctgggtGGGTGGTGTCAAAGTATTG3'

Lower case sequences add Att sites for Gateway cloning.

4)VDL1 or VDL2 KD, acetyl CoA carboxylase promoter and terminator, NAT on the same transcript.



Primers to amplify antisense regions:

VDL1, Fwd: 5'ggggacaagtttgtacaaaaaagcaggctCAGTTGTCCAAATCCTCGCTCC3'

VDL1, Rev: 5'ggggaccactttgtacaagaaagctgggtaACAGTCGTCCTAGGCACCATTG3'

VDL2, Fwd: 5'ggggacaagtttgtacaaaaaagcaggctCCCAAACACTTGGCAGTACACG3'

VDL2, Rev: 5'ggggaccactttgtacaagaaagctgggtaTCAATGCCTACTCGGTCGATAC3'

Lower case sequences add Att sites for Gateway cloning.

CHAPTER 3

Enhanced triacylglycerol (TAG) and protein accumulation in transgenic diatom *Thalassiosira pseudonana* with altered photosynthetic pigmentation

3.1 ABSTRACT

Microalgal productivity in mass cultures is limited by the inefficiency with which available light energy is utilized for photochemistry. In dense cultures, cells closest to the light source absorb more light energy than they can use and dissipate the excess while light penetrance into the culture is steeply attenuated. Reducing microalgal light harvesting and/or dissipating capacity may improve the efficiency with which light is utilized by mass cultures. In this study, two transgenic lines of the diatom *Thalassiosira pseudonana* with altered photopigmentation are evaluated with respect to photosynthetic parameters, growth, and productivity. In one line, violaxanthin de-epoxidase-like 2 is overexpressed (VDL2 OE), resulting in a reduction of the diadinoxanthin cycle pigments, which are involved in light energy dissipation (non-photochemical quenching, NPQ), accompanied by a stoichiometric increase in the light-harvesting pigment fucoxanthin. No differences in the maximum potential quantum yield of photosystem II (Fv/Fm) or light-limited photosynthetic rate (α) were found. However, when adapted to 30 µmol photons m^{-2} sec^{-1}, the VDL2 OE maximum relative electron transport rate ($rETR_{max}$) upon exposure to saturating light intensities was 86-95% of wild type (WT). When adapted to 300 µmol photons m^{-2} sec^{-1}, VDL2 OE saturated photosynthesis at 62-71% of the light intensity needed to saturate WT (E_k). NPQ was substantially lower at and below 300 µmol photons m^{-2} sec^{-1}. VDL2 OE accumulated up to 3.4 times as much triacylglycerol (TAG) as WT during exponential growth, and up to twice as much protein. Growth was up to 7% slower, but harvesting could be timed to achieve improved yields. TAG and protein accumulation inversely correlated with NPQ. The second strain evaluated was obtained by using antisense to simultaneously silence or knock down (KD) both LUT1-like genes, hypothesized to catalyze an intermediate carotenoid biosynthesis step of converting β-carotene to zeaxanthin. Overall reduction of photopigment content without altering photopigment ratios resulted. No significant differences from WT in photosynthetic performance were found. LTL KD grew at a rate comparable

to WT and accumulated up to 40% more TAG during exponential growth, while protein content was reduced by 11-19%. LTL KD cells were elongated and 5-10% smaller than WT, and cultures contained auxospores, indicating stress that may relate to a cell cycle progression defect.

3.2 INTRODUCTION

Microalgae are a promising production platform for sustainable biofuels and other bioproducts. They do not compete with plant crops for arable land and require only modest nutrient input and light energy to synthesize biomolecules of commercial interest and accumulate biomass [Mata et al. 2010]. A major challenge for the economic feasibility of microalgal production is the inefficient utilization of light by microalgal cultures grown at scale [De Mooij et al. 2015]. Microalgae have evolved extensive photopigmentation, which confers a competitive advantage in the wild by maximizing a cell's light absorption capacity in environments where light may be limited, while minimizing the light available to cells below. In dense cultures such as those used for production, this results in a suboptimal distribution of light energy. At high light intensities such as direct sunlight, algae closest to the light source capture more light than the cells are able to utilize. Excess light is dissipated as heat and fluorescence. Because of the efficiency of light capture, there is a steep attenuation of light penetrance into the culture. Thus, the cells closest to the light source are subject to photosystem-damaging light-induced stress, while the cells deeper in the culture have less light available for photosynthesis [De Mooij et al. 2015]. In theory, reducing cellular light-harvesting and/or dissipation capacity could improve light distribution and therefore productivity in dense cultures, making their cultivation for commercial purposes more cost-efficient. Cells closest to the light source would suffer less light-induced stress and light penetrance would increase. Thus, a greater proportion of the culture would be photosynthetically active [De Mooij et al. 2015].

Numerous efforts have been made in chlorophytes to reduce the size of photoantennae, which serve to capture light energy and funnel it to the photosynthetic reaction centers where it is used to drive photochemistry. This has generally resulted in improved photosynthetic parameters, such as saturation of photosynthesis at higher irradiances, greater light-saturated rates of oxygen evolution on a per-chlorophyll basis [Beckmann et al. 2009, Cazzaniga et al. 2014, Jeong et al. 2017, Kirst et al. 2012a, Kirst et al. 2012b, Mitra and Melis 2008, Nakajima et al. 2001, Shin et al. 2016, Shin et al. 2017], increased quantum yield and lower photoinhibition [Mussgnug et al. 2007]. In some cases, higher maximal culture density [Polle at. al. 2003], faster growth [Mussgnug et al. 2007], and better biomass productivity in laboratory conditions have been reported [Beckman et al. 2009, Shin et al. 2016, Shin et al. 2017]. Several strains that appeared promising based on laboratory performance did not show improved biomass productivity in mass culture conditions simulated in laboratory-scale panel photobioreactors, possibly due to unintended effects of the genetic modifications or higher vulnerability to photodamage [De Mooij et al. 2015]. Cazzaniga et al. [2014], however, reported improved biomass productivity in laboratory conditions as well as in 7 L hanging bag photobioreactors deployed outdoors. As a different strategy, Berteotti et al. [2016] explored downregulation of light energy dissipation through non-photochemical quenching (NPQ) in the chlorophyte *Chlamydomonas reinhardtii*, and found that if it is reduced, but not completely abolished, improved biomass productivity in a small scale photobioreactor results.

Non-chlorophyte eukaryotic microalgae have been largely unexplored with respect to light utilization efficiency improvement through biological modification. Because microalgae are incredibly diverse and have evolved different strategies for interacting with their environment, different taxa may respond to such modifications with varying degrees of success. Diatoms, for example, are brown microalgae belonging to the Stramenopile or heterokont class, whose light-harvesting and photoprotective strategies differ substantially from chlorophytes [Wilhelm et al.

2006]. Chlorophytes rely predominantly on chlorophylls a and b (Chl a, Chl b) for light capture, the ratio of which adjusts dynamically in response to changes in light intensity. Diatoms use Chl a in association with chlorophyll c (Chl c) and the more abundant carotenoid-derived accessory photopigment fucoxanthin (Fx) to capture light energy [Wilhelm et al. 2006]. The ratio of Chl a to Fx, and thus photoantenna size, does not change much with light intensity [Lepetit et al. 2012]. Diatoms appear to rely on their capacity to induce NPQ faster and to a higher extent than chlorophytes when adjusting to short-term irradiance increases and coordinately reduce the abundance of photosynthetic reaction centers and photoantennae during long-term adaptation to higher light [Lepetit et al. 2012]. NPQ in diatoms relies predominantly on the diadinoxanthin (Ddx) cycle that is absent in chlorophytes, wherein the carotenoid derivative Ddx is reversibly converted to diatoxanthin (Dtx) via de-epoxidation when it is necessary to dissipate excess light energy [Lepetit et al. 2012, Wilhelm et al. 2006]. Because diatoms are very promising in terms of biomass productivity and triacylglycerol (TAG, neutral lipid of interest for biofuels) accumulation [Hildebrand et al. 2012], it is intriguing to explore improving their productivity by modulating culture light utilization efficiency. So far, only one such study has been published. A *Cyclotella* sp. strain was obtained through two subsequent rounds of mutagenesis, employing ethylmethylsulfonate and ultraviolet radiation [Huesemann et al. 2009]. Its green color indicated that it had drastically reduced carotenoid abundance, including the main accessory light-harvesting photopigment Fx and the photoprotective Ddx cycle pigments. It also had a substantially higher Chl a/Chl c ratio, indicating a smaller photoantenna size. The mutant required higher light intensity to saturate photosynthesis on a per chlorophyll basis but was less stable in culture and had reduced biomass productivity. It was not fully characterized, but the observed lack of fitness could be attributed to too much reduction in carotenoids and thus susceptibility to photodamage and oxidative stress, and

possible additional undesirable mutations [Huesemann et al. 2009]. More exploration is necessary to determine the utility of reducing light absorption and/or dissipation in diatoms.

In this study, we evaluate photosynthetic parameters, growth, carbon partitioning, and macromolecule accumulation in two transgenic (TG) lines of the diatom *Thalassiosira pseudonana* in which the abundance of two carotenoid biosynthesis enzymes is manipulated (Chapter 2). In one line, VDL2 OE, the violaxanthin de-epoxidase 2 is overexpressed, resulting in a decrease of the photoprotective Ddx cycle pigments (Ddx+Dtx) and a stoichiometric increase in the light-harvesting Fx. In the other line, LTL KD, both copies of LUT1-like, hypothesized to convert β-carotene to zeaxanthin, are simultaneously knocked down, resulting in an overall reduction of total cellular photopigment content (Tot) with conserved ratios of individual pigments. Four clones of each TG line were selected for characterization and compared to wild type (WT) in low light (LL, 30 µmol photons m^{-2} sec^{-1}) and high light (HL, 300 µmol photons m^{-2} sec^{-1}). There were four experimental culture sets: VDL2 OE vs. WT in LL, VDL2 OE vs. WT in HL, LTL KD vs. WT in LL, and LTL KD vs. WT in HL. Each set was independently acclimated to cultivation conditions and used to obtain samples and data, which were then processed independently of each other. Thus, our findings will be discussed as comparisons within but not between the sets.

3.3 RESULTS

3.3.1 Photosynthetic Parameters

Photosynthetic parameters were determined for LL and HL-adapted cultures via rapid light-response curves (RLCs) obtained with a pulse amplitude modulation (PAM) fluorometer. This approach involves measuring dynamic rather than steady state responses of a culture to gradual increases in light intensity, and allows, therefore, an assessment of the photosynthetic performance of a culture as it relates to the light intensity it had previously adapted to [Malapascua et al. 2014,

Ralph and Gaderman 2005]. The parameters assessed were Fv/Fm (maximum potential quantum yield of photosystem II), α (initial slope of the RLC curve, light-limited photosynthetic rate), $rETR_{max}$ (maximum light-saturated relative electron transport rate), E_k ($rETR_{max}/\alpha$, minimum saturating light intensity), and NPQ (non-photochemical quenching, dissipation of photon energy) [Malapascua et al. 2014, Ralph and Gaderman 2005].

No significant differences in Fv/Fm were found between the TG lines and WT in either LL or HL **(Table 3.1)**. α, $rETR_{max}$, and E_k were derived from rETR measurements over a range of irradiance, depicted in **Fig. 3.1**. LL-adapted VDL2 OE clones had a reduced $rETR_{max}$, calculated to be 86-95% of the WT average (p-value = 0.02) **(Fig. 3.1A, Table 3.1)**. Nevertheless, rETR values did not vary between the TG lines and WT in any tested condition when measured at the irradiance they were adapted to **(Fig. 3.1)**. Additionally, HL-adapted VDL2 OE clones had reduced E_k values, calculated to saturate photosynthesis at 62-71% of the average WT minimum saturating light intensity (p-value = 0.008) **(Fig. 3.1B, Table 3.1)**. Although some trends in parameter differences between the TG lines and WT could be observed, no other significant differences were found.

NPQ values measured over a range of irradiance are depicted in **Fig. 3.2**. A high degree of variability between different TG clones and WT cultures was observed, and no claims about statistically significant differences consistent between all TG clones within any condition could thus be made, with one exception. At lower irradiances, the HL-adapted VDL2 OE clones had substantially less NPQ than WT **(Fig. 3.2B)**. At 125 µmol photons m^{-2} sec^{-1}, VDL2 OE clone 6 had 53% of the average WT NPQ, while the other three clones were at 1-8% of the WT average (p-value = 0.01). At 191 µmol photons m^{-2} sec^{-1}, VDL2 OE clone 6 measured at 64% of the WT average NPQ, while the other three clones had 16-18% of the WT average (p-value = 0.02). At 282 µmol photons m^{-2} sec^{-1}, VDL2 OE clone 6 had 82% of the average WT NPQ, while the other three clones were at 27-36% (p-value = 0.08 with clone 6, and 0.03 without).

3.3.2 Morphology, Cell, and Chloroplast Dimensions

The VDL2 OE clones appeared morphologically similar to WT (data not shown). LTL KD clones, on the other hand, were elongated and formed numerous auxospores, indicating sexual reproduction **(Fig. 3.3A, B)**. The elongation phenotype and auxospore formation were more prominent in HL, but observable in LL as well. Incubating with PDMPO, a fluorescent dye that incorporates into newly synthesized silica, confirmed that the elongated cells were single cells, not multiple cells that have failed to separate, as no cell wall could be observed within the cells **(Fig. 3.3C, D)**. The LTL KD cells with a high degree of elongation appeared to have numerous chloroplasts distributed throughout **(Fig. 3.3E, F)**.

The average cell area was not significantly different between WT and VDL2 OE clones but was reduced in LTL KD clones in HL, measuring at 90-95% of the WT average (p-value = 0.03) **(Table 3.2)**. The chloroplast to cell area ratio in LTL KD clones did not significantly differ from WT in LL or HL **(Table 3.2)**. In HL-adapted VDL2 OE clones, the chloroplast to cell area ratio was reduced to 93-96% of the WT average (p-value = 0.02).

3.3.3 Growth Rates and Maximal Culture Density

During exponential growth in LL, the specific growth rate of the VDL2 OE clones was approximately 98% of the WT average (p-value = 0.0006) **(Table 3.2)**. When light-limiting culture densities were reached around day 6, the disparity between the KD clones and WT cultures became more pronounced **(Fig. 3.4A)**. The stationary phase culture densities reached by the VDL2 OE clones in LL were 72-88% of the WT average (p = 0.02).

In HL, VDL2 OE clones also experienced slowing **(Fig. 3.4B)**. During exponential growth, the specific growth rate of the VDL2 OE clones was 93-96% of the WT average (p-value = 0.009) **(Table**

3.2). At stationary phase, HL clones VDL2 OE3 and OE5 reached 54% of the average WT culture density, OE2 was at 84% of WT, and OE6 did not differ from WT. Interestingly, all the HL-adapted VDL2 OE clones were substantially lower in culture density than WT during day 5 (34-57% of the WT average, p-value = 0.002) and day 6 (45-69% of the WT average, p-value = 0.007), as the light-limited cultures were transitioning to stationary phase.

The LTL KD clones had some slowing in growth compared to WT in LL and in HL **(Fig. 3.4C, D)**. Unlike the VDL2 OE clones that had the most substantial slowing during the light-limited transition from exponential to stationary phase, the slowing in LTL KD clones was more uniform, from inoculation to stationary phase **(Fig. 3.4)**. The specific growth rates of the LTL KD clones in LL were 97-99% of the WT average (p-value = 0.04), and not significantly different in HL **(Table 3.2)**. Stationary phase culture densities of the LTL KD clones did not significantly differ from WT.

3.3.4 Lipid, Protein, Carbohydrate, and Photopigment Content

The cellular abundance of neutral lipids (triacylglycerol, TAG, quantified by BODIPY fluorescence), proteins, carbohydrates, and photopigments was assessed during exponential growth.

The TAG content of VDL2 OE clones was found to be 21-81% greater than the WT average (p-value = 0.05) in LL, and 2-3.4 times greater than the WT average in HL (p-value = 0.04) **(Fig. 3.5A, B)**. The LTL KD TAG content did not significantly differ from WT in LL. In HL, LTL KD clone 49 had TAG content similar to WT, while the other 3 clones had 25-40% more **(Fig. 3.5C, D)**. When normalized to average cell area, TAG content was 17-46% greater than the WT average in HL (p-value = 0.03) **(Fig. S3.1A)**.

No significant differences were found when comparing total cellular protein between TG lines and WT in LL **(Fig. 3.6A, C)**. HL-adapted VDL2 OE clones contained approximately 1.5-2 times as much protein as WT (p-value = 0.02) **(Fig. 3.6B)**. HL-adapted LTL KD clones had 81-89% of the average WT total cellular protein content (p-value = 0.01) **(Fig. 3.6D)**. When normalized to average cell area, total cellular protein content did not significantly differ between HL-adapted LTL KD clones and WT **(Fig. S3.1B)**.

No significant differences in total carbohydrate content were found between TG lines and WT, in either LL or HL **(Fig. S3.2)**. Cell area-normalized carbohydrate content also did not differ significantly between LTL KD clones and WT in HL **(Fig. S3.1C)**.

Average cell area (3.3.2) and total cellular photopigment content **(Chapter 2)** did not significantly differ between VDL2 OE clones and WT. LL-adapted LTL KD clones had 75-93% total photopigments of the WT average (p-value = 0.2), and 60-76% of the average WT total photopigments in HL (p-value = 0.009) **(Chapter 2)**. In HL, cell area-normalized total photopigment content was reduced in LTL KD clones to 63-81% of the WT average (p-value = 0.02).

3.3.5 Inverse Relationship Between NPQ and TAG, Protein Content

Plotting NPQ measured at the light intensity closest to the irradiance to which the cultures were adapted (125 µmol photons m^{-2} sec^{-1} for LL, 282 µmol photons m^{-2} sec^{-1} for HL) **(Fig. 3.2)** against BODIPY fluorescence **(Fig. 3.5)**, revealed an inverse trend between the two variables for VDL2 OE clones and WT in LL and HL **(Fig. 3.7)**. The Pearson correlation coefficient (R), where 0 signifies no correlation and -1 means that there is a perfect negative correlation, was -0.3 in LL and -0.4 in HL. No such trend could be observed for LTL KD at LL, as there was no significant variance in TAG or NPQ in that condition, nor for LTL KD in HL, as there was no significant variance in NPQ **(Figs. 3.2C, D, 3.5C)**.

Significant difference from WT in protein content was found only for VDL2 OE in HL **(Fig. 3.6B)**. When plotted against NPQ as described above, an inverse trend, stronger than that for TAG, was observed as well (R = -0.7) **(Fig. 3.8A)**. LL-adapted LTL KD clone 51 had approximately 40% less cell area-normalized protein than the average of the other LTL KD clones and WT cultures **(Fig. 3.6C)**. This correlated with approximately 1.9 times more NPQ than the average of the other LTL KD clones and WT cultures at the measured value closest to the irradiance to which the cultures were adapted, **(Fig. 3.2C)**, thus confirming the trend of an inverse relationship between NPQ and protein accumulation (R = -0.9) **(Fig. 3.8B)**.

3.4 DISCUSSION

3.4.1 Photosynthesis

In VDL2 OE clones, Tot did not exhibit a trend with respect to WT, and the ratios of Chl a, Chl c, and β-carotene to Tot were unchanged **(Chapter 2)**. This indicates that the stoichiometric increase in Fx/Tot at the expense of (Ddx+Dtx)/Tot did not affect the cellular abundance of photosynthetic reaction centers and photoantennae. Rather, Fx had replaced some of what would have been Ddx cycle pigments, which are present in distinct pools. Some are bound to photoantenna proteins, while others are dissolved in the lipid shield around photoantennae [Lepetit et al. 2010]. Without further investigation outside the scope of this study, it is not possible to directly discern which pool(s) of the Ddx cycle pigments were at least partially replaced with Fx in the VDL2 OE clones.

Photosynthesis in VDL2 OE clones was not negatively impacted in LL or HL, as indicated by no change in Fv/Fm, α, and rETR measured at the irradiance to which the cultures were adapted with respect to WT **(Table 3.1, Fig. 3.1A, B)**. The lowered $rETR_{max}$ in LL-adapted VDL2 OE clones

means that in the short term, LL-adapted VDL2 OE clones would have an impairment in photosynthetic electron transfer through photosystem II compared to WT upon being exposed to saturating irradiance **(Table 3.1, Fig. 3.1A)** [Malapascua et al. 2014, Ralph and Gaderman 2005]. This is not the case with HL-adapted VDL2 OE clones, the $rETR_{max}$ of which did not significantly differ from WT **(Table 3.1, Fig. 3.1B)**. Photosynthetic electron transport of the HL-adapted VDL2 OE clones, however, would, in the short term, saturate at a lower irradiance than in WT, as indicated by lower E_k values **(Table 3.1, Fig. 3.1B)** [Malapascua et al. 2014, Ralph and Gaderman 2005]. The mechanism by which altered photopigmentation in VDL2 OE clones causes the observed defects in short-term adaptation to higher irradiance needs to be investigated further. It may be attributable to the excess, misplaced Fx.

As discussed in Chapter 2, the overall photopigment reduction in LTL KD clones is most likely due to the destabilization of light-harvesting complexes caused by Fx insufficiency, as carotenoids are known to be crucial for their assembly and stabilization [Moskalenko and Karapetyan 1996, Santabarbara et al. 2013]. Nevertheless, in comparison to WT, photosynthetic parameters as indicated by Fv/Fm, α, and rETR at the irradiance to which the cultures were adapted were not adversely affected in the LTL KD clones in LL or HL **(Table 3.1, Fig. 3.1C, D)**. In contrast to the VDL2 OE clones, short-term adaptation to increases in irradiance was also not impaired, as demonstrated by $rETR_{max}$ and E_k values not differing between the LTL KD clones and WT **(Table 3.1, Fig. 3.1)**.

The high degree of variance between different TG clones in NPQ responses over a range of irradiance, especially in the LL-adapted state in which WT cultures had a highly replicable response **(Fig. 3.2)**, may be explained by potential inter-clonal differences in adaptation to altered photopigmentation. The substantially reduced NPQ in HL-adapted VDL2 OE lines at lower irradiance

levels **(Fig. 3.2B)** may be attributed to the reduced Ddx cycle pigment abundance, as the largest part of NPQ in diatoms requires the presence of Dtx [Lepetit et al. 2012]. The extent of NPQ induction depends on the light adaptation state and incident irradiance, and it appears that despite the reduced abundance of Ddx cycle pigments, HL-adapted VDL2 OE clones were able to induce WT-equivalent NPQ levels at higher irradiance levels **(Fig. 3.2B)** [Lepetit et al. 2012].

3.4.2 Auxospore Formation in LTL KD

Frequent auxospore formation and cell elongation, as observed for the LTL KD clones **(Fig. 3.3)**, are both markers of stress. Under typical laboratory conditions, auxospore formation, which indicates sexual reproduction, is rarely if ever observed in *T. pseudonana* [Moore et al. 2017]. In some diatom species, sexual reproduction serves as a way to reconstitute cell size, as it diminishes with every division. However, *T. pseudonana* appears to maintain a relatively constant cell size. Sexual reproduction in diatoms may also be triggered by growth stress that leads to cell cycle arrest, such as nutrient depletion and oxidative stress [Moore et al. 2017]. Because the LTL KD clones were grown in the same nutrient-replete media as WT, which did not form abundant auxospores, nutrient depletion is not a likely cause of the observed sexual reproduction in LTL KD clones. Oxidative stress is also unlikely, since it would be expected to lead to reduced Fv/Fm, which was not observed **(Table 3.1)**. A small portion of LTL KD cultures consisted of cells that were smaller than typical (data not shown), and it is possible that sexual reproduction was induced in those cells to restore size. However, stress evidenced by elongation **(Fig. 3.3)** was present in the majority of LTL KD cells, and we suggest that it was a contributor to auxospore formation. The elongated phenotype has been documented in *T. pseudonana* subject to various stresses that impede cell cycle progression, such as copper toxicity and limitation in silica or selenium. It has not been observed in response to nitrate or phosphorus limitation and is thus not a universal stress response

[Davis et al. 2005]. The cause of the stress in our study is not clear. We hypothesize that it may relate to a defect in chloroplast division, stemming from light-harvesting assembly impairment and destabilization by Fx deficiency in LTL KD clones. Microalgae coordinate cell and chloroplast division to ensure that both daughter cells have chloroplasts upon cytokinesis [Sumiya et al. 2016]. Depending on the timing, an arrest in chloroplast division may cause cell cycle arrest, or chloroplasts may continue dividing without concomitant cell cycle progression if cells are arrested in S-phase, resulting in numerous chloroplasts per cell [Sumiya et al. 2016]. A deregulated coordination between chloroplast division and cell cycle progression in the LTL KD clones could thus account for cell elongation and auxospore formation, which may result from a cell cycle progression defect, as well as the overaccumulation of chloroplasts observed in some of the highly elongated cells **(Fig. 3.3)**.

3.4.3 Growth, Carbon Partitioning, and Macromolecule Accumulation

The most disparity in growth rates between VDL2 OE clones and WT was observed during the light-limited linear portions of the growth curves in LL and HL **(Fig. 3.4A, B)**. This may be explained by excess Fx reducing light availability or energy transfer efficiency in the culture when light became limiting. The disparity lessened at higher culture densities, likely because at that point the difference in shading experienced by VDL2 OE clones and WT cultures diminished.

Exponential growth rates for VDL2 OE clones were up to 7% lower than in WT in HL and 2% lower in LL, and average cell area was not statistically different **(Table 3.2)**. However, VDL2 OE clones accumulated substantially more TAG per cell than WT in LL and HL **(Fig. 3.5A, B)**, and HL-adapted VDL2 OE clones also accumulated more protein per cell than WT **(Fig. 3.6B)**. This suggests that VDL2 OE clones may have been fixing more carbon than WT, and preferentially storing it as

TAG (and producing more protein in HL), rather than using it to fuel faster growth or storing it as carbohydrate **(Fig. S3.2B)**. Thus, excess fixed carbon was diverted to glycolysis, which eventually feeds into TAG and amino acid biosynthesis, rather than gluconeogenesis, which results in carbohydrate biosynthesis [Smith et al. 2012].

NPQ dissipates absorbed light energy and is inversely correlated with the amount of photons available for photochemistry. Because NPQ was measured as part of an RLC curve on dark-adapted cultures rather than in real time during cultivation, the measurements at the light intensity closest to what cultures had been adapted to (cultivation irradiance) are estimates of the NPQ response in cultures. We observed an inverse correlation between NPQ at the cultivation irradiance and TAG levels for VDL2 OE clones and WT cultures in LL and HL **(Fig. 3.7)**. The most striking difference in TAG accumulation as well as in NPQ closest to cultivation irradiance as compared to WT was measured for HL-adapted VDL2 OE clones **(Figs. 3.2B, 3.5B)**. HL-adapted VDL2 OE clones were also the only condition that accumulated significantly more protein per cell than WT **(Fig. 3.6B)**, and an inverse relationship between cellular protein content and NPQ closest to cultivation irradiance was also found for that condition **(Fig. 3.8A)**. In LL, LTL KD clone 51 had substantially less protein per cell than the other clones and WT cultures, and this correlated with substantially higher NPQ as well **(Figs. 3.2C, 3.6C, 3.8B)**. Additionally, that clone exhibited markedly slower growth in LL than the other cultures **(Fig. 3.4C)**. We suggest that diminished NPQ in cultivation conditions allowed the VDL2 OE clones to utilize more photons for carbon fixation, which was then used to synthesize additional TAG and protein. Conversely, excess NPQ at cultivation irradiance for LL-adapted LTL KD clone 51 would have reduced the amount of light energy available for photosynthesis, resulting in slower growth and reduced protein content. For VDL2 OE clones and WT at LL and HL, extent of NPQ measured closest to irradiance correlated with (Ddx+Dtx)/Tot (R = 0.7 in LL and 0.8 in HL, where 0 signifies no correlation and 1 a perfect positive correlation) **(Fig.**

S3.3). No such correlation was found for LTL KD clone 51, and the increased magnitude of NPQ it exhibited compared to the other LTL KD clones may be attributed to a clonal difference in adapting the light-harvesting machinery to reduced photopigmentation.

HL-adapted LTL KD clones had higher TAG than WT **(Figs. 3.5D, S3.1A)**, but no significant variation in NPQ. We suggest that in this case, enhanced TAG accumulation occurred due to growth stress (3.4.2), with some of the fixed carbon stored as TAG instead of being used to fuel growth due to a defect in cell cycle progression. The observed reduced abundance of proteins and photopigments per cell area and smaller cell size could result from such stress as well. Additionally, improved light penetrance into the culture due to reduced cellular pigmentation in LTL KD clones may have contributed to increased carbon fixation and storage as TAG. The milder phenotype observed for LTL KD clones in LL may be explained by a lesser extent of photopigment reduction with respect to WT than in HL.

3.4.4 Concluding Remarks

Our results indicate that reducing the photoprotective Ddx cycle pigments without reducing light-harvesting pigmentation may be a promising strategy for improving TAG and protein productivity in diatoms. Ddx cycle pigments are necessary for NPQ induction and preventing photoinhibition due to excess absorbed light energy and resultant oxidative stress [Lepetit et al. 2010]. However, diatoms may accumulate them in excess of what is needed to protect cells without unnecessarily reducing the amount of light energy available for photochemistry. A similar concept has been reported by Berteotti et al. [2016], who found that downregulating but not completely abolishing NPQ in the chlorophyte *Chlamydomonas reinhardtii* improved biomass productivity in laboratory conditions.

As detailed in Chapter 1, strain performance in laboratory conditions does not always relate to productivity in a production setting. Therefore, it will be important to assess the performance of VDL2 OE in production conditions and explore diel changes in TAG and protein abundance under sinusoidal light and temperature encountered outdoors, as well as productivity throughout the growth curve, to find optimal harvesting conditions. Our preliminary data are promising: in HL, which was closer to production conditions that typically utilize even higher irradiance, harvesting could be timed so as to obtain improved TAG yields, despite the slight growth rate reduction in VDL2 OE clones. For example, VDL2 OE clone 6 grew at the same rate as WT prior to reaching a light-limiting density, while accumulating more TAG and protein **(Figs. 3.4B, 3.5B, 3.6B)**. Thus, if harvested prior to slowing, VDL2 OE clone 6 would yield approximately 3-fold more TAG and 2-fold more protein **(Fig. 3.5B, 3.6B)**. Other clones were slower than WT throughout the growth curve **(Fig. 3.4B)** but could be harvested to yield an equivalent or increased amount of TAG and protein compared to WT harvested a day later, for example, reducing operating costs. The enhanced protein content in VDL2 OE clones could be used as a high-value co-product for applications such as animal feed, further offsetting the costs of TAG production [Moreno-Garcia et al. 2017]. Although it is not yet known how VDL2 OE clones will perform in a production setting, it is encouraging that Cazzaniga et al. [2014] found that their *Chlorella sorokiniana* mutant with reduced photoantenna size had better biomass productivity than WT outdoors as well as in laboratory conditions.

The ability to enhance TAG production without adversely affecting growth is an important goal for advancing biofuels, as currently their commercialization is stymied by production inefficiency [Trentacoste et al. 2013]. Our VDL2 OE clones suffered somewhat from slower growth compared to WT, especially in light-limiting culture densities. This may be at least partially attributed to excess Fx reducing the light utilization efficiency in cultures. Reducing Ddx cycle pigments without increasing the amount of Fx may result in an improvement over VDL2 OE by

reducing NPQ without impairing growth. As described in Chapters 1 and 2, the abundance of Ddx cycle pigments increases with cultivation irradiance. Some Ddx cycle pigments are bound to photoantenna proteins, while others, including the majority of those synthesized in response to increased irradiance, are dissolved in the lipid shield that surrounds photoantennae [Lepetit et al. 2012]. Light-induced biosynthesis of Ddx cycle pigments appears differentially regulated from the induction of carotenoid biosynthesis when photoantenna proteins need to be populated during chloroplast division (Chapter 1). In Chapter 2, we hypothesized that the two copies of phytoene synthase (PSY, catalyzes the first committed step of carotenoid biosynthesis) found in the *T. pseudonana* genome may serve to differentially activate carotenoid biosynthesis in response to different cellular needs, with PSY1 serving during chloroplast division and PSY2 during irradiance increase. Thus, knocking down PSY2 may be a promising strategy for reducing Ddx cycle pigments without affecting Fx content. It may prove especially useful in sinusoidal light, such as encountered outdoors. As described in Chapter 1, Ddx cycle pigment abundance closely follows light intensity changes when a sinusoidal light regime is applied. Reducing the amplitude of that response by knocking down PSY2, if it functions as hypothesized, may improve light utilization efficiency throughout the photoperiod under sinusoidal light. Another useful approach may be to target proteins involved in NPQ, such as LHCX3 [Hao et al. 2018].

Reducing overall photopigment content may also be considered for TAG productivity improvement in diatoms. Although the LTL KD clones had a slightly reduced protein content and did not accumulate as much additional TAG as VDL2 OE clones did, there was an approximately 25-40% improvement over WT in three out of the four clones **(Fig. 3.5D)**. The LTL KD clones had an advantage over VDL2 OE clones in that their growth was not substantially different from WT, even when light-limited **(Fig. 3.4)**. The disadvantage of the LTL KD clones was the apparent stress experienced by the cells, which may potentially lead to reduced culture stability in production

conditions. Nevertheless, it will be interesting to evaluate the performance of LTL KD clones in production conditions to assess the utility of overall photosynthetic pigment reduction for improving productivity in diatoms.

3.5 METHODS

3.5.1 Cultivation, Sampling, and Growth Curves

WT and TG *T. pseudonana* cultures were cultivated at either 30 or 300 µmol photons m^{-2} sec^{-1} (natural white LED lighting, superbrightleds.com, NFLS-NW300X3-WHT-LC2), using a 12:12 light:dark regime, at 18°C. 50 mL cultures in Erlenmeyer flasks were maintained in Artificial Sea Water (ASW) medium [Darley and Volcani 1969] with rapid stirring. Each experimental set included 2 WT cultures and 4 TG clones. After inoculation, the cultures were grown to 1-3x10^6 cells/mL, then allowed to adapt to the cultivation conditions by daily dilutions that maintained exponential growth with culture density under 2.5x10^6 cells/mL for a minimum of 2 weeks prior to sampling. Cultures were rotated between stir plates each day to minimize any position-specific differences and transferred to clean flasks once a week. Sampling for protein content, carbohydrate content, cell/chloroplast dimensions and lipid content, and photosynthetic measurements was performed on separate days, within the first two hours of the light period. After sampling was completed, the cultures were inoculated into fresh 50 mL of ASW for growth curves at approximately 8-11x10^3 cells/mL. Cell counts were performed in triplicate daily, including immediately upon inoculation, with the MUSE® Cell Analyzer (EMD Millipore, Billerica, MA) and averaged. Specific growth rates were calculated as $\ln(x_1-x_0)/t$, where x_1 = number of cells at the end of exponential growth, x_0 = number of cells at the beginning of exponential growth, t = number of exponential growth days.

3.5.2 Photosynthetic Parameter Measurements

Photosynthetic parameter measurements were performed using a Walz WATER-PAM fluorometer (Heinz Walz GbmH, Eichenring, Germany). The fluorometer was calibrated by using ASW as a blank. Cultures at 1-3x10^6 cells/mL were dark-adapted for at least 45 minutes prior to measurements. Measurements were performed on 3-4 replicate aliquots from the same culture stocks. Fv/Fm, α, $rETR_{max}$, and E_k were calculated by the WinControl-3 software (Heinz Walz GbmH, Eichenring, Germany).

3.5.3 Silica Staining and Microscopy

PDMPO (2-(4-pyridyl)-5-((4-(2-dimethylaminoethylaminocarbamoyl)methoxy)phenyl)oxazole), a fluorescent dye that binds to freshly incorporated silica [Shimizu et al. 2001], was used to monitor cell wall formation. 5 mL aliquots of exponentially growing cultures were incubated with .125 μM PDMPO in 40 mL glass culture tubes under the cultivation conditions for 24 hours, allowing for approximately two cell doublings. Cells were imaged using a Zeiss Axio Observer Z1 Inverted Microscope (Carl Zeiss Microimaging Inc., USA). The Zeiss#05 (Ex 395–440 nm, FT 460 nm, Em 470nm LP) filter was used for chlorophyll autofluorescence and Zeiss #21HE (Ex 387/15 nm, FT 409, Em 510/90 nm) was used for PDMPO. Images were acquired using a 40x objective and processed with the AxioVision 4.7.2 software (Carl Zeiss Microimaging Inc., USA).

3.5.4 Cell/Chloroplast Dimensions and BODIPY Fluorescence

2.5-8.3x10^7 exponentially growing cells per sample were harvested by centrifugation and stored at -20°C until processing. Pellets were thawed on ice and resuspended in 0.5 mL 2.3% NaCl. 1.3 μL of 1 mg/mL stock of the lipophilic fluorescent dye BODIPY (4,4-difluoro-4-bora-3a,4a-diaza-s-

indacene, 493/503, Molecular Probes, ThermoFisher Scientific, USA) per sample was added to stain for neutral lipids (TAG). Following a 30 min incubation on ice in the dark, data for 10,000 cells per sample were collected using an ImageStream X imaging flow cytometer with the INSPIRE™ software package (Amnis Corp., Seattle, WA, USA). 0.6 and 1.0 neutral density filters and 488 nm excitation were used during data aquisition. Post-acquisition spectral compensation and data analysis were performed with the IDEAS™ software (Amnis Corp., Seattle, WA, USA) [Hildebrand et al. 2015]. After debris, unfocused cells, and images containing more than one cell were discarded, 1200-7200 cells were analyzed per sample.

3.5.5 Protein Content

1.5-7.2x10^7 cells per replicate were harvested by centrifugation and stored at -20°C until processing. Pellets were thawed on ice and resuspended in 6x volume of extraction buffer (4% SDS, 125 mM Tris-Cl pH 6.8), incubated at 95°C for 5 min, then centrifuged at maximum speed for 3 min. Supernatants were transferred to clean tubes. Protein concentrations were measured using the DC™ Protein Assay (Bio-Rad, Hercules, CA, USA), based on the Lowry method for protein quantification, according to manufacturer's instructions. Absorbance measurements at 750 nm were performed in triplicate using a SpectraMax M2 microplate reader (Molecular Devices LLC, San Jose, CA, USA). A bovine gamma globulin standard (Bio-Rad, Hercules, CA, USA) set was used to generate a standard curve for calculating protein concentrations in the samples.

3.5.6 Carbohydrate Content

Total cellular carbohydrate content was determined using a method adapted from Granum and Myklestad [2002]. 1.8-6.3x10^7 cells per replicate were harvested by centrifugation, washed in

2.3% NaCl, and stored at -20°C until processing. Pellets were thawed on ice, resuspended in 1 mL 0.05 M H_2SO_4, incubated in a 60°C water bath for 10 min, then centrifuged at 4000 g for 2 min. Two 400 µL supernatant aliquots per sample were transferred to clean 1.5 mL microfuge tubes for duplicate analysis. 100 µL of 3% freshly prepared aqueous phenol and 1 mL concentrated H_2SO_4 were added to each tube, followed by vortexing. After a 30 min incubation at room temperature, samples absorbance at 485 nm was measured in a 1 cm quartz cuvette using a DU™ 730 spectrophotometer (Beckman Coulter, Brea, CA, USA). 155, 38.75, 9.68, and 2.42 mg/L glucose solutions were used to generate a standard curve for calculating carbohydrate concentrations in the samples.

3.5.7 Statistical Analysis

An online one-way ANOVA calculator (https://www.socscistatistics.com/tests/anova/default2.aspx) was used to assess the statistical significance of the difference or lack thereof between measurements obtained for TG lines and WT cultures. An online Pearson correlation coefficient calculator (https://www.socscistatistics.com/tests/pearson/Default2.aspx) was used to assess two-variable correlations.

3.6 ACKNOWLEDGEMENTS

Chapter 3, in full, is material currently being prepared for submission for publication. Gaidarenko, Olga; Yee, Daniel; Hildebrand, Mark. “Enhanced triacylglycerol (TAG) and protein accumulation in transgenic diatom *Thalassiosira pseudonana* with altered photosynthetic pigmentation.” Olga Gaidarenko was the principal researcher and author of this work. We thank Dr. Andrew E. Allen and members of his lab for providing access to the Water PAM used in this work.

We also thank Ms. Corinne Sathoff for assistance with cultivation chamber set-up and Image Stream data analysis. Additionally, we thank Dr. James Golden for critically reading this work and providing helpful input. This work was supported by U.S. Dept. of Energy grant DE-FOA-0001471.

Table 3.1.

Photosynthetic parameters. VDL2 overexpression (OE) and LTL knockdown (KD) clones are compared to wild-type (WT) cultures in low light (30 µmol photons m^{-2} sec^{-1}, LL) and high light (300 µmol photons m^{-2} sec^{-1}, HL). Fv/Fm = maximum quantum yield of photosystem II (PSII), α = light-limited photosynthetic rate, $rETR_{max}$ = maximum relative electron transport rate, E_k = minimum saturating irradiance.

		Fv/Fm	α	$rETR_{max}$	E_k
VDL2 OE LL	WT1	0.698 ± 0.009	0.214 ± 0.032	119.0 ± 6.3	570.3 ± 99.7
	WT2	0.707 ± 0.007	0.185 ± 0.021	118.3 ± 6.7	652.3 ± 112.6
	OE2	0.697 ± 0.005	0.166 ± 0.006	112.3 ± 5.6	676.9 ± 34.9
	OE3	0.714 ± 0.030	0.158 ± 0.010	102.1 ± 9.2	653.7 ± 96.1
	OE6	0.713 ± 0.003	0.191 ± 0.031	104.3 ± 10.8	570.3 ± 162.5
	OE7	0.688 ± 0.003	0.196 ± 0.002	102.9 ± 11.7	524.5 ± 64.1
	p-value	*0.9*	*0.3*	*0.02*	*0.9*
VDL2 OE HL	WT1	0.699 ± 0.008	0.158 ± 0.008	217.1 ± 36.3	1368.6 ± 179.4
	WT2	0.679 ± 0.007	0.135 ± 0.000	151.0 ± 5.9	1119.9 ± 49.0
	OE2	0.685 ± 0.008	0.175 ± 0.019	146.9 ± 7.8	850.5 ± 113.7
	OE3	0.680 ± 0.002	0.173 ± 0.026	133.7 ± 16.5	803.6 ± 194.3
	OE6	0.678 ± 0.005	0.156 ± 0.024	115.8 ± 10.4	770.8 ± 166.4
	OE7	0.685 ± 0.008	0.164 ± 0.031	136.4 ± 19.4	882.8 ± 257.9
	p-value	*0.4*	*0.1*	*0.09*	*0.008*
LTL KD LL	WT1	0.711 ± 0.009	0.184 ± 0.018	125.4 ± 8.2	692.9 ± 95.9
	WT2	0.713 ± 0.007	0.210 ± 0.007	119.9 ± 10.3	573.5 ± 67.1
	KD49	0.693 ± 0.009	0.194 ± 0.022	110.7 ± 3.9	578.6 ± 68.4
	KD51	0.703 ± 0.011	0.190 ± 0.010	104.5 ± 4.3	550.7 ± 33.8
	KD60	0.706 ± 0.007	0.197 ± 0.020	119.9 ± 5.8	617.0 ± 95.4
	KD71	0.709 ± 0.009	0.183 ± 0.017	117.3 ± 8.9	650.3 ± 111.5
	p-value	*0.2*	*0.6*	*0.2*	*0.5*
LTL KD HL	WT1	0.639 ± 0.006	0.141 ± 0.007	97.9 ± 17.9	702.8 ± 150.5
	WT2	0.625 ± 0.002	0.147 ± 0.003	85.0 ± 5.0	581.4 ± 44.2
	KD49	0.609 ± 0.004	0.158 ± 0.008	81.8 ± 11.2	517.4 ± 55.4
	KD51	0.628 ± 0.009	0.153 ± 0.016	72.2 ± 11.0	483.1 ± 113.4
	KD60	0.625 ± 0.004	0.172 ± 0.023	92.3 ± 10.5	554.6 ± 131.9
	KD71	0.647 ± 0.027	0.154 ± 0.014	92.9 ± 15.1	617.0 ± 138.9
	p-value	*0.7*	*0.08*	*0.5*	*0.2*

Table 3.2.

Average cell and chloroplast (Chl) area measurements, their ratio, and specific growth rates. VDL2 overexpression (OE) and LTL knockdown (KD) clones are compared to wild-type (WT) cultures in low light (30 µmol photons m^{-2} sec^{-1}, LL) and high light (300 µmol photons m^{-2} sec^{-1}, HL).

		Cell Area	Chl Area	Chl/Cell Area	Specific Growth Rate (µ, day^{-1})
VDL2 OE LL	WT1	37.8 ± 4.8	22.7 ± 4.5	0.60	2.26
	WT2	39.4 ± 5.0	23.7 ± 4.5	0.60	2.27
	OE2	40.2 ± 5.8	23.7 ± 5.1	0.59	2.22
	OE3	39.0 ± 6.3	22.1 ± 5.5	0.57	2.22
	OE6	37.1 ± 5.8	19.5 ± 5.5	0.53	2.22
	OE7	37.3 ± 6.0	21.4 ± 5.2	0.57	2.21
	p-value	*0.9*	*0.3*	*0.1*	*0.0006*
VDL2 OE HL	WT1	38.2 ± 5.9	23.1 ± 4.7	0.60	3.86
	WT2	39.5 ± 5.8	23.1 ± 5.0	0.59	3.85
	OE2	38.0 ± 7.0	21.0 ± 6.8	0.55	3.62
	OE3	36.5 ± 6.9	20.6 ± 6.5	0.56	3.71
	OE6	38.2 ± 6.3	21.6 ± 6.1	0.56	3.58
	OE7	39.8 ± 6.8	22.7 ± 6.8	0.57	3.68
	p-value	*0.5*	*0.08*	*0.02*	*0.009*
LTL KD LL	WT1	41.6 ± 5.8	20.7 ± 5.7	0.50	1.83
	WT2	43.3 ± 6.5	22.8 ± 6.1	0.53	1.83
	KD49	40.4 ± 6.7	23.9 ± 6.7	0.59	1.79
	KD51	41.2 ± 6.8	24.5 ± 7.5	0.59	1.77
	KD60	40.8 ± 7.1	24.6 ± 7.7	0.60	1.80
	KD71	39.5 ± 6.7	20.4 ± 6.1	0.52	1.81
	p-value	*0.06*	*0.4*	*0.1*	*0.04*
LTL KD HL	WT1	48.4 ± 7.4	24.9 ± 6.2	0.51	3.80
	WT2	48.0 ± 8.3	26.0 ± 6.6	0.54	3.72
	KD49	43.2 ± 6.9	22.3 ± 5.1	0.52	3.68
	KD51	45.3 ± 8.3	24.2 ± 6.4	0.53	3.59
	KD60	45.5 ± 8.3	24.0 ± 7.0	0.53	3.66
	KD71	46.0 ± 8.7	23.1 ± 7.0	0.50	3.54
	p-value	*0.03*	*0.05*	*0.6*	*0.06*

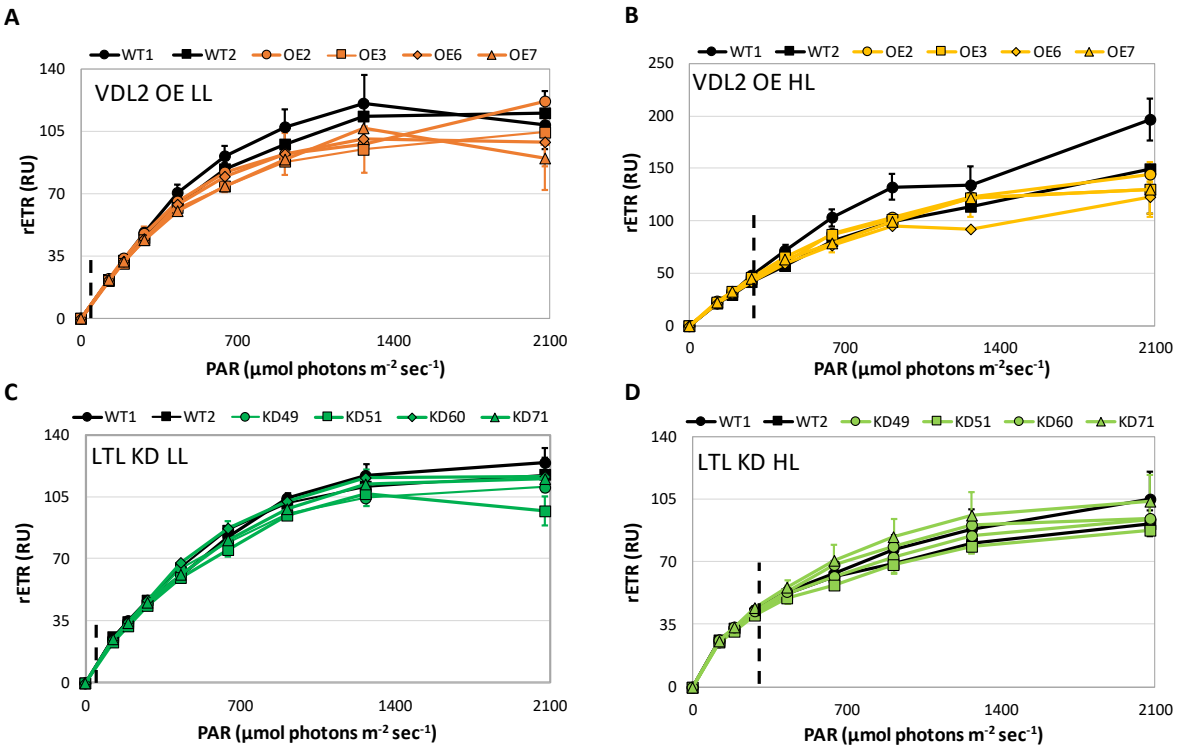


Figure. 3.1.

Rapid light curves, relative electron transfer rates (rETR – relative units, RU) vs. photosynthetically active radiation (PAR). Vertical dashed black lines indicate irradiance to which the cultures were adapted. Data are presented as averages of 3-4 replicates ± standard deviation. **A.** 30 µmol photons m^{-2} sec^{-1} (low light - LL)-adapted VDL2 overexpression (OE) clones vs. wild-type (WT); **B.** 300 µmol photons m^{-2} sec^{-1} (high light - HL)-adapted VDL2 OE clones vs. WT; **C.** LL-adapted LTL knockdown (KD) clones vs. WT; **D.** HL-adapted LTL KD clones vs. WT.

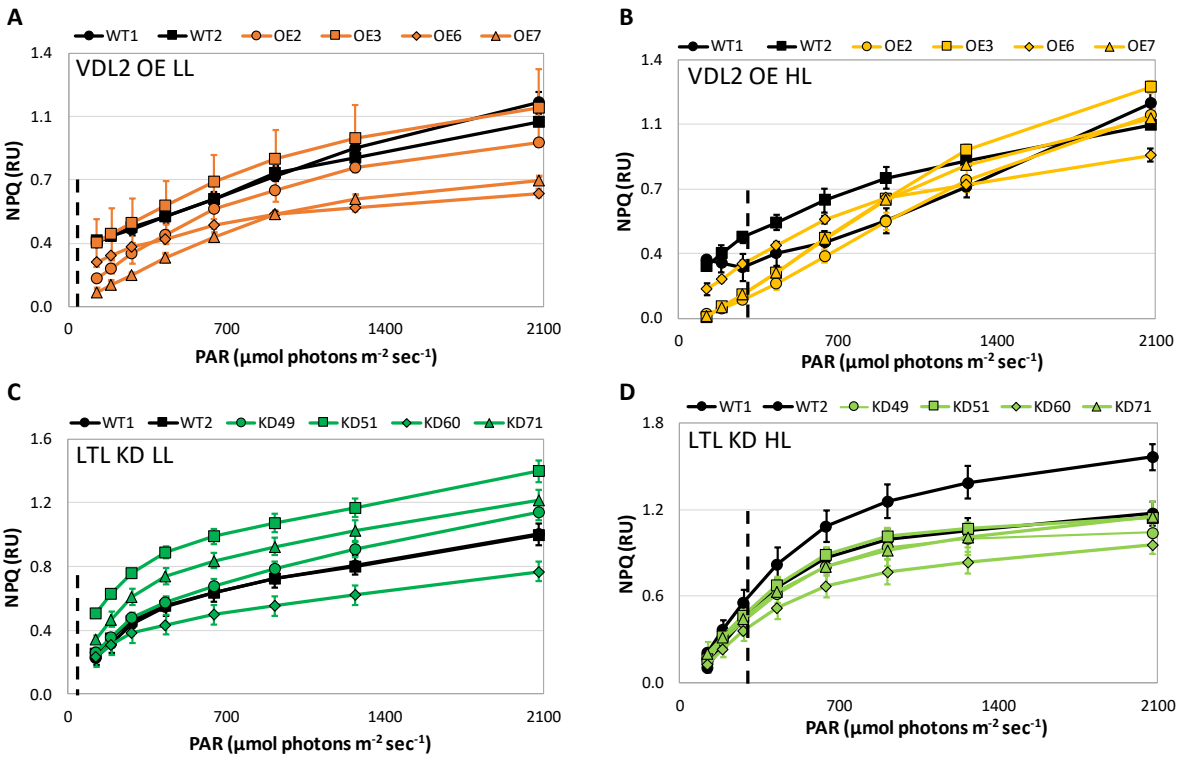


Figure. 3.2.

Rapid light curve-derived non-photosynthetic quenching values (NPQ – relative units, RU) vs. photosynthetically active radiation (PAR). Vertical dashed black lines indicate irradiance to which the cultures were adapted. Data are presented as averages of 3-4 replicates ± standard deviation. **A.** 30 µmol photons m^{-2} sec^{-1} (low light - LL)-adapted VDL2 overexpression (OE) clones vs. wild-type (WT); **B.** 300 µmol photons m^{-2} sec^{-1} (high light - HL)-adapted VDL2 OE clones vs. WT; **C.** LL-adapted LTL knockdown (KD) clones vs. WT; **D.** HL-adapted LTL KD clones vs. WT.

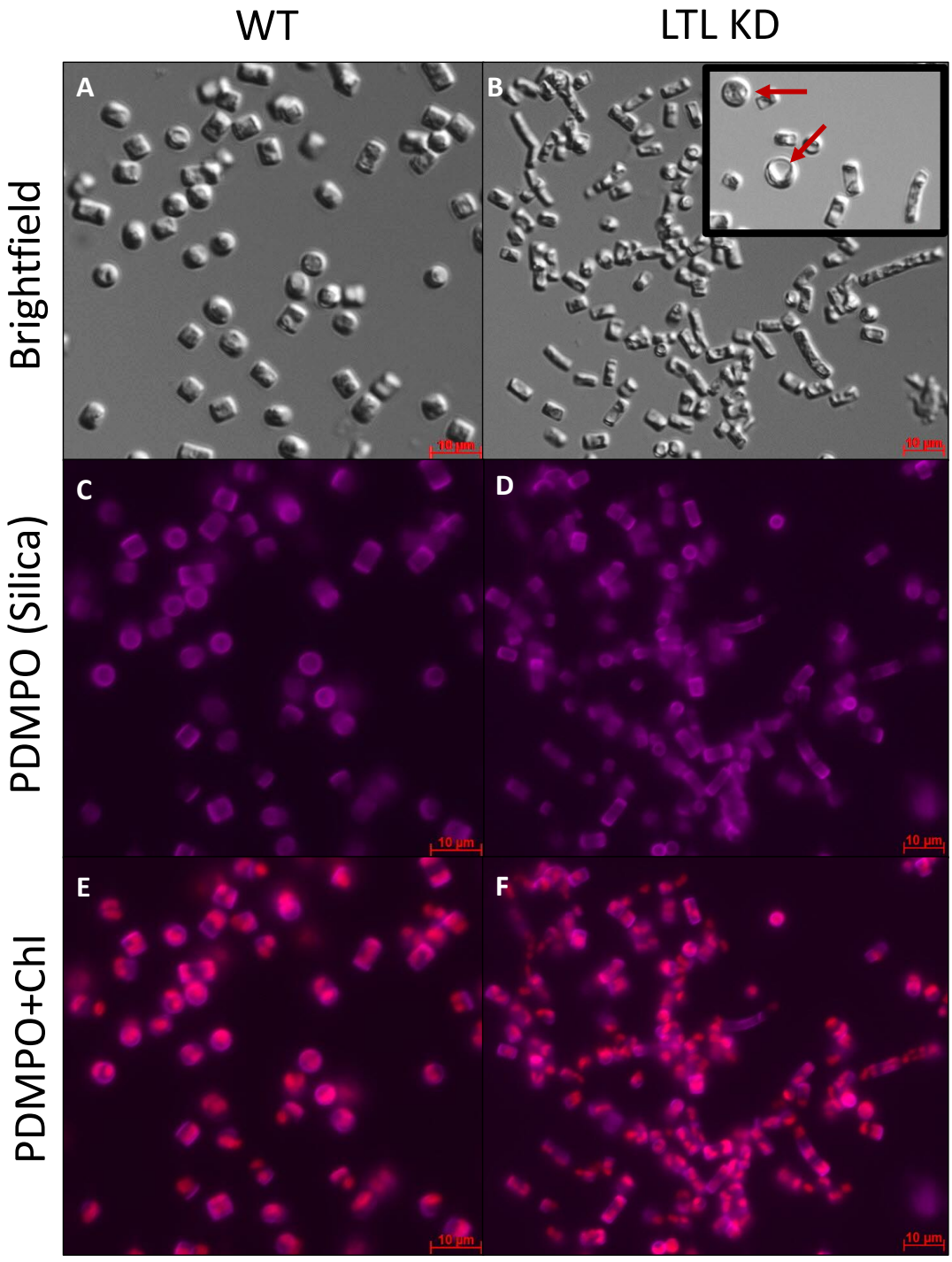


Figure. 3.3.
Wild-type (WT) (A, C, E) and LTL knockdown (KD) (B, D, F) cultures, cultivated at 300 µmol photons m^{-2} sec^{-1} (high light - HL). **A, B.** Brightfield. Red arrows indicate auxospores; **C, D.** PDMPO staining for silica; **E, F.** PDMPO and chlorophyll fluorescence (Chl).

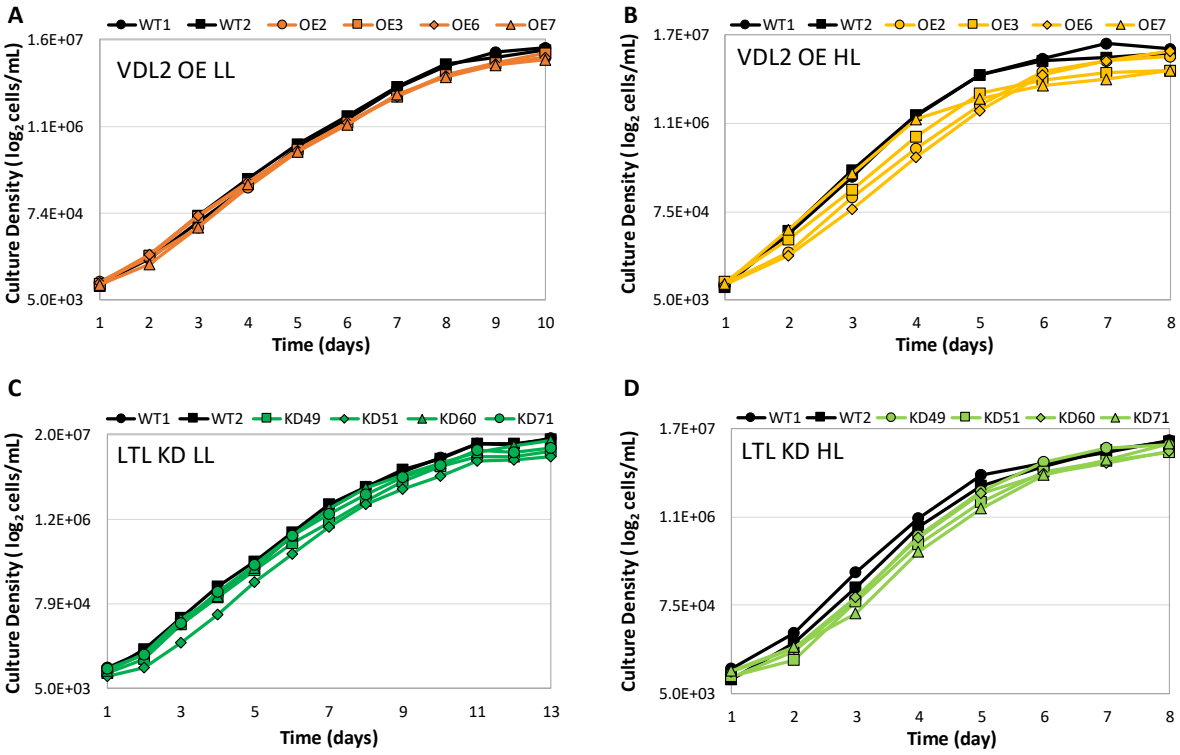


Figure. 3.4.

Growth curves. Each data point represents an average of three technical replicates.

A. 30 µmol photons m^{-2} sec^{-1} (low light - LL)-adapted VDL2 overexpression (OE) clones vs. wild-type (WT); **B.** 300 µmol photons m^{-2} sec^{-1} (high light - HL)-adapted VDL2 OE clones vs. WT; **C.** LL-adapted LTL knockdown (KD) clones vs. WT; **D.** HL-adapted LTL KD clones vs. WT.

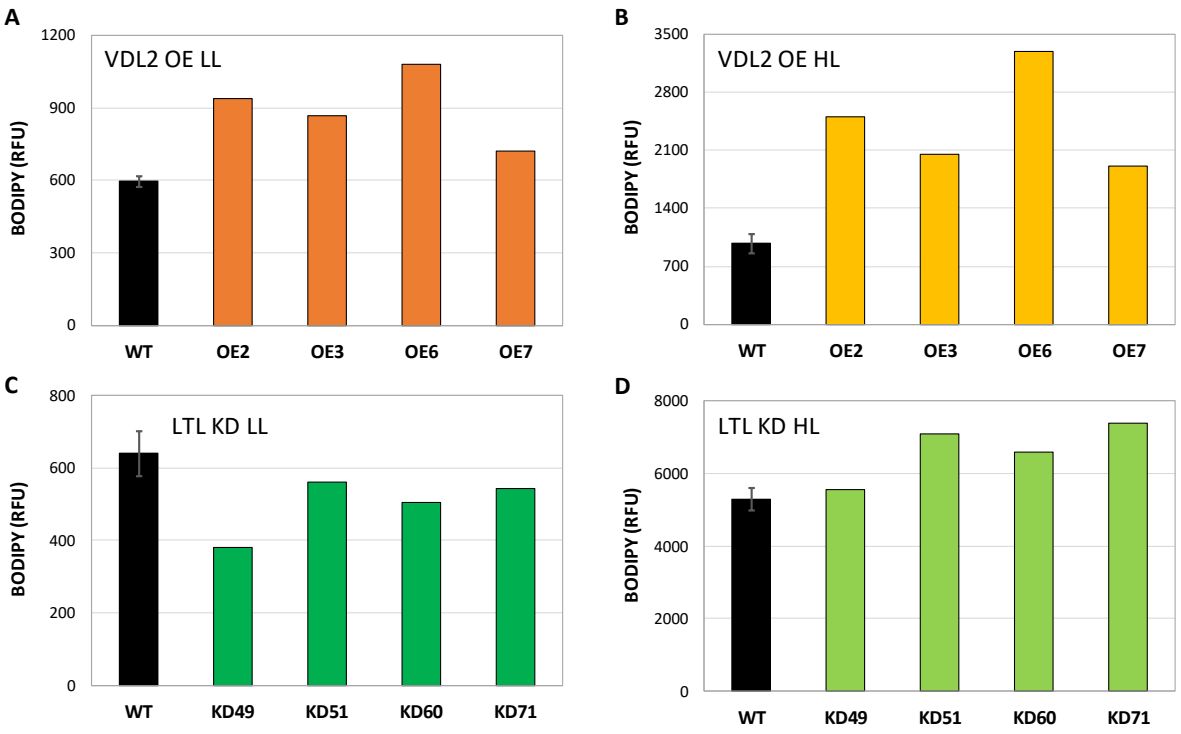


Figure. 3.5.

Average BODIPY fluorescence (relative fluorescence units – RFU). Wild-type (WT) data are presented as averages of two independent cultures ± standard deviation. **A.** 30 µmol photons m^{-2} sec^{-1} (low light - LL)-adapted VDL2 overexpression (OE) clones vs. wild-type (WT); **B.** 300 µmol photons m^{-2} sec^{-1} (high light - HL)-adapted VDL2 OE clones vs. WT; **C.** LL-adapted LTL knockdown (KD) clones vs. WT; **D.** HL-adapted LTL KD clones vs. WT.

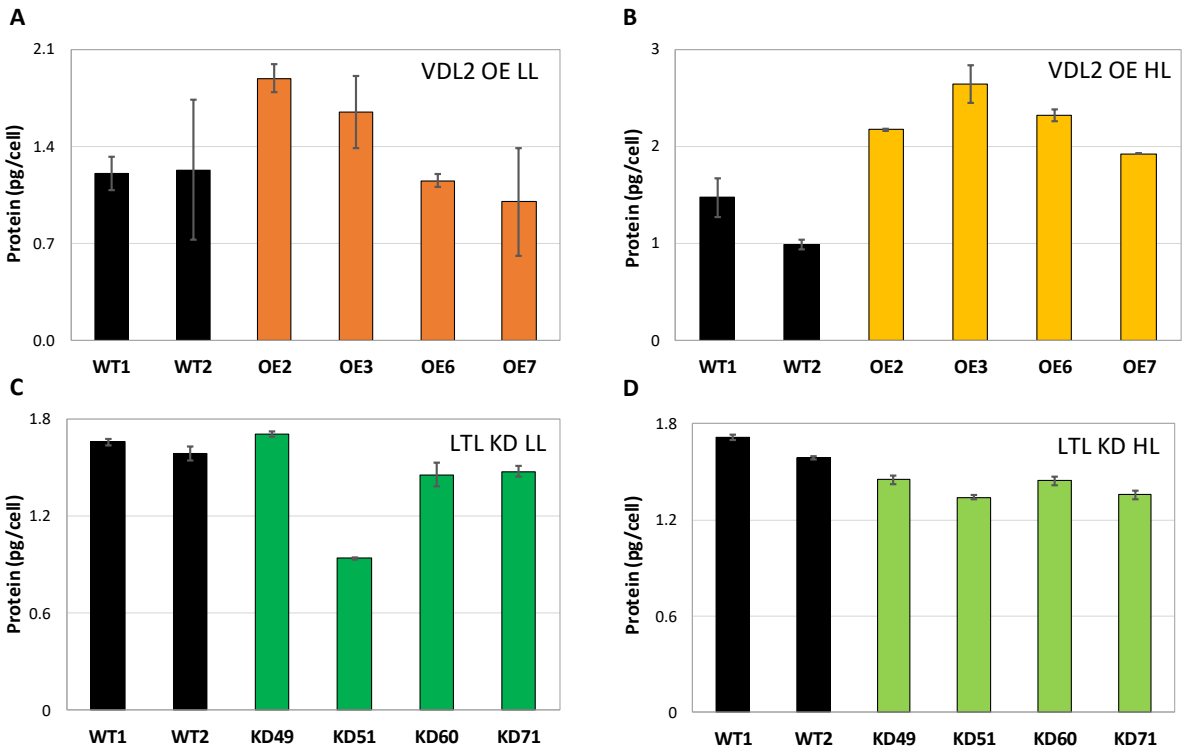


Figure. 3.6.

Average total cellular protein content. Data are presented as averages of two technical replicates ± standard deviation. **A.** 30 µmol photons m^{-2} sec^{-1} (low light - LL)-adapted VDL2 overexpression (OE) clones vs. wild-type (WT); **B.** 300 µmol photons m^{-2} sec^{-1} (high light - HL)-adapted VDL2 OE clones vs. WT; **C.** LL-adapted LTL knockdown (KD) clones vs. WT; **D.** HL-adapted LTL KD clones vs. WT.

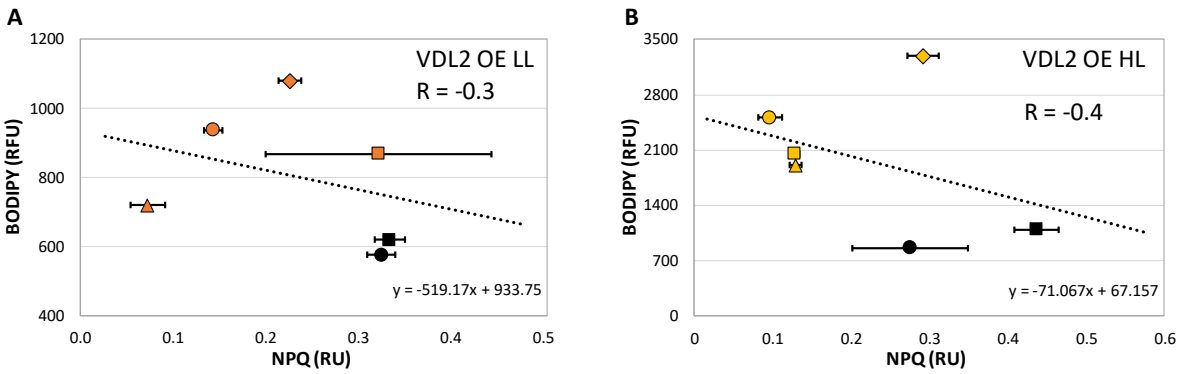


Figure. 3.7.
Relationship between non-photochemical quenching (NPQ, relative units - RU) at irradiance to which the cultures were adapted and average BODIPY fluorescence (relative fluorescence units, RFU). Data are presented as averages of 3-4 technical replicates ± standard deviation for NPQ. Pearson correlation coefficient (R) is indicated on the plots. Wild-type (WT) cultures are represented by black circles (WT1) and squares (WT2). VDL2 overexpression (OE) clones are represented by orange circles (VDL2 OE2), squares (VDL2 OE3), rhombuses (VDL2 OE6), and triangles (VDL 2 OE7). **A.** 30 µmol photons m^{-2} sec^{-1} (low light - LL)-adapted VDL2 OE clones vs. WT; **B.** 300 µmol photons m^{-2} sec^{-1} (high light - HL)-adapted VDL2 OE clones vs. WT.

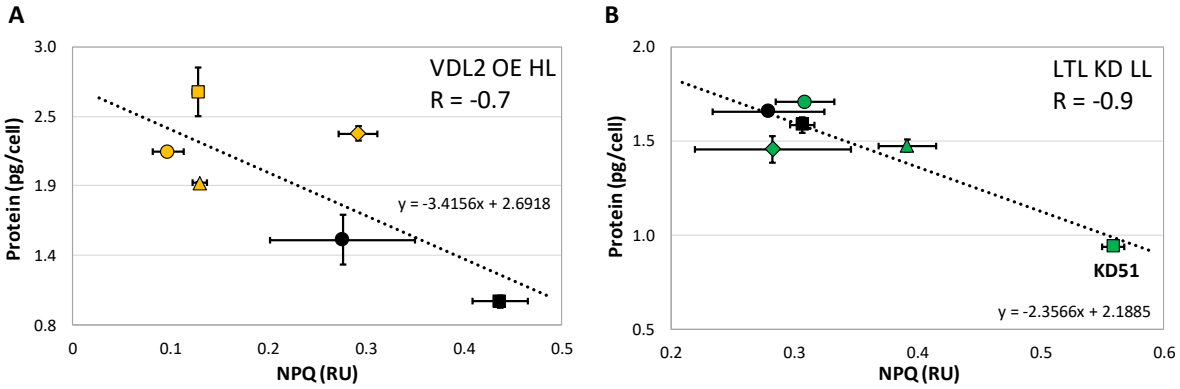


Figure. 3.8.
Relationship between non-photochemical quenching (NPQ – relative units, RU) at irradiance to which the cultures were adapted and average cellular protein content. Data are presented as averages of two technical replicates ± SD for protein content, and as averages of 3-4 technical replicates ± SD for NPQ. Pearson correlation coefficient (R) is indicated on the plots. Wild-type (WT) cultures are represented by black circles (WT1) and squares (WT2). VDL2 overexpression (OE) clones are represented by orange circles (VDL2 OE2), squares (VDL2 OE3), rhombuses (VDL2 OE6), and triangles (VDL 2 OE7). LTL knockdown (KD) clones are represented by green circles (LTL KD49), squares (LTL KD51), rhombuses (LTL KD60), and triangles (LTL KD71). **A.** 300 µmol photons m^{-2} sec^{-1} (high light - HL)-adapted VDL2 OE clones vs. WT; **B.** 30 µmol photons m^{-2} sec^{-1} (low light - LL)-adapted LTL KD clones vs. WT. Clone LTL KD51 is labeled on the plot.

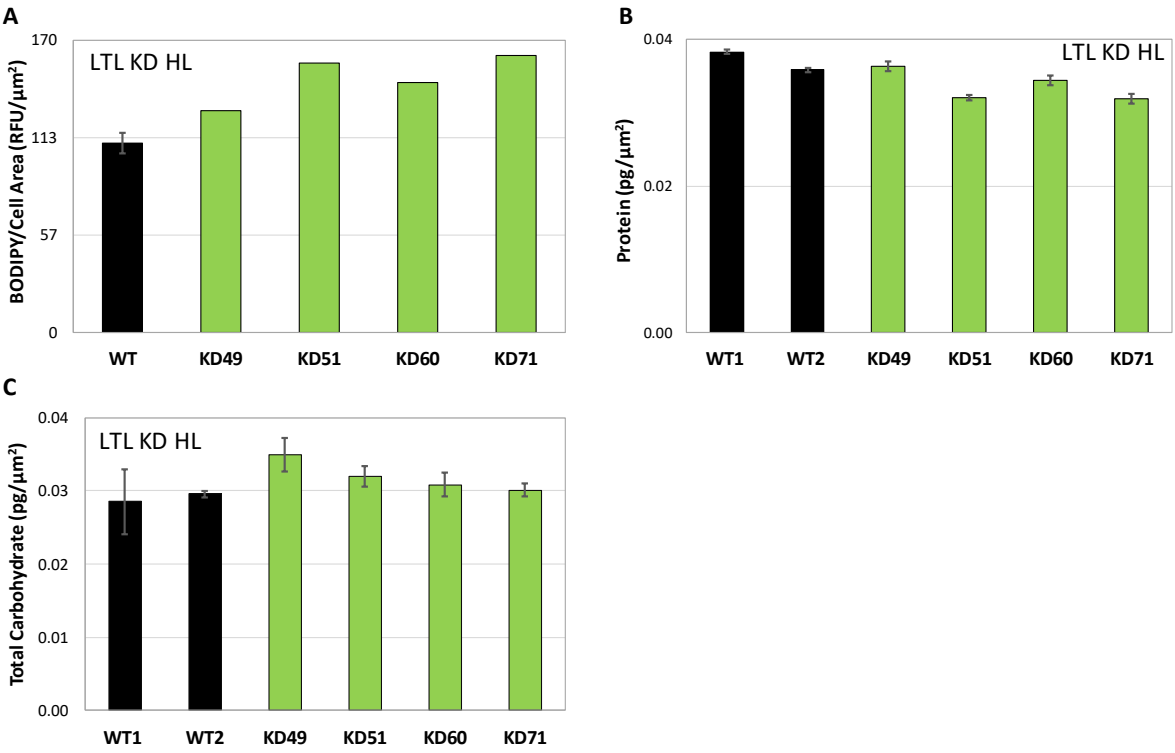


Figure. S3.1.

A. Average BODIPY fluorescence (relative fluorescence units – RFU) normalized by average cell area. 300 µmol photons m^{-2} sec^{-1} (high light - HL)-adapted LTL knockdown (KD) clones vs. wild-type (WT). WT data is presented as an average of two independent cultures ± standard deviation; **B.** Average cellular protein content normalized by average cell area. HL-adapted LTL KD clones vs. WT. Data are presented as averages of two technical replicates ± standard deviation. **C.** Average total cellular carbohydrate content normalized by average cell area. HL-adapted LTL KD clones vs. WT. Data are presented as averages of two technical replicates ± standard deviation.

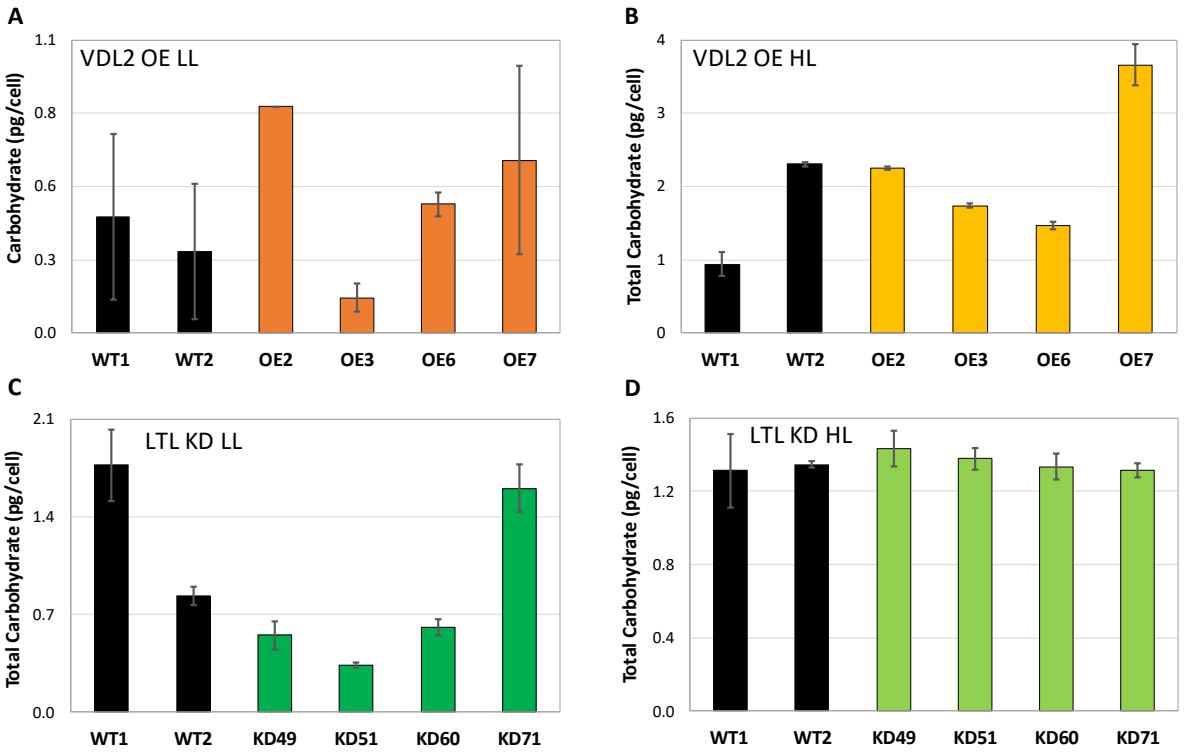


Figure. S3.2.

Average total cellular carbohydrate content. Data are presented as averages of two technical replicates ± standard deviation. **A.** 30 µmol photons m^{-2} sec^{-1} (low light - LL)-adapted VDL2 overexpression (OE) clones vs. wild-type (WT); **B.** 300 µmol photons m^{-2} sec^{-1} (high light - HL)-adapted VDL2 OE clones vs. WT; **C.** LL-adapted LTL knockdown (KD) clones vs. WT; **D.** HL-adapted LTL KD clones vs. WT.

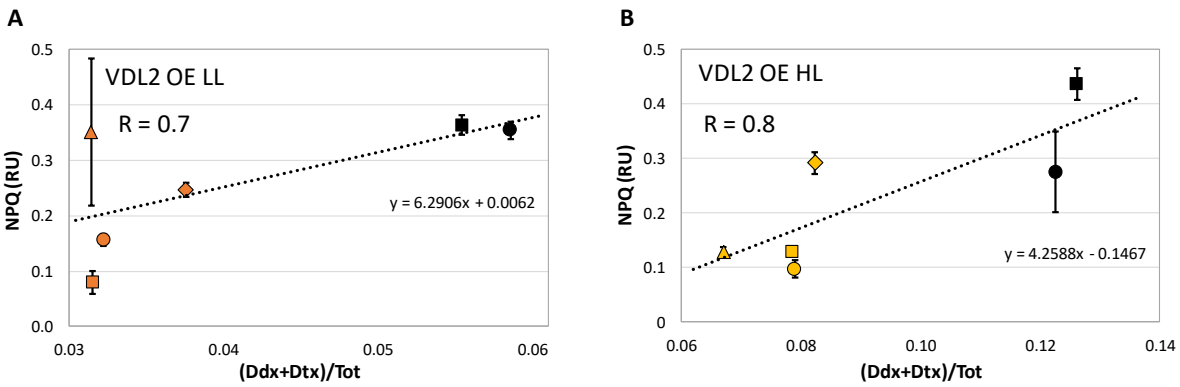


Figure. S3.3.
Relationship between the ratio of diadinoxanthin cycle pigments to total cellular photopigment content ((Ddx+Dtx)/Tot) and non-photochemical quenching (NPQ, relative units - RU) at irradiance to which the cultures were adapted. Data are presented as averages of 3-4 technical replicates ± standard deviation for NPQ. Pearson correlation coefficient (R) is indicated on the plots. Wild-type (WT) cultures are represented by black circles (WT1) and squares (WT2). VDL2 overexpression (OE) clones are represented by orange circles (VDL2 OE2), squares (VDL2 OE3), rhombuses (VDL2 OE6), and triangles (VDL 2 OE7). **A.** 30 µmol photons m^{-2} sec^{-1} (low light - LL)-adapted VDL2 OE clones vs. WT; **B.** 300 µmol photons m^{-2} sec^{-1} (high light - HL)-adapted VDL2 OE clones vs. WT.

3.7 REFERENCES

Beckmann J., Lehr F., Finazzi G., Hankamer B., Posten C., Wobbe L., Kruse O. (2009) Improvement of light to biomass conversion by de-regulation of light-harvesting protein translation in *Chlamydomonas reinhardtii*. J Biotechnol 142: 70-77.

Berteotti S., Ballottari M., Bassi R. (2016) Increased biomass productivity in green algae by tuning non-photochemical quenching. Sci Rep 6:21339.

Cazzaniga S., Dall'Osto L., Szaub J., Scibilia L., Ballotari M., Purton S., Bassi R. (2014) Domestication of the green alga *Chlorella sorokiniana*: reduction of antenna size improves light-use efficiency in a photobioreactor. Biotechnol Biofuels 7:157.

Darley W., Volcani B. (1969) Role of silicon in diatom metabolism: a silicon requirement for deoxyribonucleic acid synthesis in the diatom *Cylindrotheca fusiformis* Reimann and Lewin. Exp Cell Res 58: 334–42.

Davis A. K., Hildebrand M., Palenik B. (2005) A stress-induced protein associated with the girdle band region of the diatom *Thalassiosira pseudonana* (Bacillariophyta). J Phycol 41: 577-589.

de Mooij T., Janssen M., Cerezo-Chinarro O., Mussgnug J. H., Kruse O., Ballottari M., Bassi R., Bujaldon S., Wollman F., Wijffels R. H. (2015) Antenna size reduction as a strategy to increase biomass productivity: a great potential not yet realized. J Appl Phycol 27: 1063-1077.

Granum E., Myklestad S. M. (2002) A simple combined method for determination of β-1,3-glucan and cell wall polysaccharides in diatoms. Hydrobiologia 477: 155-161.

Hao T., Jiang T., Dong H., Ou L., He X., Yang Y. (2018) Light-harvesting protein Lhcx3 is essential for high light acclimation of *Phaeodactylum tricornutum.* AMB Expr 8: 174.

Hildebrand M., Davis A., Abbriano R., Pugsley H. R., Traller J. C., Smith S. R., Shrestha R., Cook O., Sanchez-Alvarez E. L., Manandhar-Shrestha K., Alderete B. (2016) Applications of imaging flow cytometry for microalgae. In: Barteneva N., Vorobjev I. (eds.), Imaging Flow Cytometry. Methods in Molecular Biology vol. 1389. Humana Press, New York, NY.

Hildebrand M., Davis A., Smith S. R., Traller J. C., Abbriano R. (2012) The place of diatoms in the biofuels industry. Biofuels 3(2): 221-240.

Huesemann M. A., Hausmann T. S., Bartha R., Aksoy M., Weissman J. C., Benemann J. R. (2009) Biomass productivities in wild type and mutant of *Cyclotella* (sp.) (diatom). Appl Biochem Biotechnol 157: 507-526.

Jeong J., Baek K., Kirst H., Melis A., Jin E. (2017) Loss of CpSRP54 function leads to a truncated light-harvesting antenna size in *Chlamydomonas reinhardtii*. Biochim Biophys Acta 1858: 45-55.

Kirst H., Garcia-Cerdan J. G., Zurbriggen A., Melis A. (2012) Assembly of the light-harvesting chlorophyll antenna in the green alga *Chlamydomonas reinhardtii* requires expression of the *TLA2-CpFTSY* gene. Plant Physiol 158: 930-945.

Kirst H., Garcia-Cerdan J. G., Zurbriggen A., Ruehle T., Melis A. (2012) Truncated photosystem chlorophyll antenna size in the green microalga *Chlamydomonas reinhardtii* upon deletion of the *TLA3-CpSRP43* gene. Plant Physiol 160: 2251-2260.

Lepetit B., Goss R., Jakob T., Wilhelm C. (2012) Molecular dynamics of the diatom thylakoid membrane under different light conditions. Photosynth Res 111(1-2): 245-257.

Lepetit B., Volke D., Gilbert M., Wilhelm C., Goss R. (2010) Evidence for existence of one antenna-associated, lipid-dissolved and two protein-bound pools of diadinoxanthin cycle pigments in diatoms. Plant Physiol 154: 1906-1920.

Malapascua J. R. F., Jerez C. G., Sergejevova M., Figueroa F. L., Masojidek J. (2014) Photosynthesis monitoring to optimize growth of microalgal mass cultures: application of chlorophyll fluorescence techniques. Aquat Biol 22: 123-140.

Mata T. M., Martins A. A., Caetano N. S. (2010) Microalgae for biodiesel production and other applications: a review. Renew Sust Energ Rev 14: 217-232.

Mitra M., Melis A. (2008) Optical properties of microalgae for enhanced biofuels production. Opt Express 16(26): 21807-21820.

Moore E. R., Bullington B. S., Weisberg A. J., Jiang Y., Chang J., Halsey K. H. (2017) Morphological and transcriptomic evidence for ammonium induction of sexual reproduction in *Thalassiosira pseudonana* and other centric diatoms. PLoS One 12(7): e0181098.

Moreno-Garcia L., Adjalle K., Sarbane S., Raghavan G. S. V. (2017) Microalgae biomass production for a biorefinery system: recent advances and the way towards sustainability. Renew Sust Energ Rev 76: 493-506.

Moskalenko A. A., Karapetyan N. V. (1996) Structural role of carotenoids in photosynthetic membranes. Z Naturforsch 51c: 763-771.

Mussgnug J. H., Thomas-Hall S., Rupprecht J., Foo A., Klassen V., McDowall A., Schenk P. M., Kruse O., Hankamer B. (2007) Engineering photosynthetic light capture: impacts on improved solar energy to biomass conversion. Plant Biotechnol J 5: 802-814.

Nakajima Y., Tsuzuki M., Ueda R. (2001) Improved productivity by reduction of the content of light-harvesting pigment in *Chlamydomonas perigranulata*. J Appl Phycol 13: 95-101.

Polle J. E. W., Kanakagiri S., Melis A. (2003) *tla1*, a DNA insertional transformant of the green alga *Chlamydomonas reinhardtii* with a truncated light-harvesting chlorophyll antenna size. Planta 217: 49-59.

Ralph P. J., Gaderman R. (2005) Rapid light curves: a powerful tool to assess photosynthetic activity. Aquat Bot 82(3): 222-237.

Santabarbara S., Casazza A. P., Ali K., Economou C. K., Wannathong T., Zito F., Redding K. E., Rappaport F., Purton S. (2013) The requirement for carotenoids in the assembly and function of photosynthetic complexes in *Chlamydomonas reinhardtii*. Plant Physiol 161: 535-546.

Shimizu K., Del Amo Y., Brzezinski M. A., Stucky G. D., Morse D. E. (2001) A novel fluorescent silica tracer for biological silicification studies. Chem Biol 8: 1051-1060.

Shin W., Lee B., Jeong B., Chang Y. K., Kwon J. (2016) Truncated light-harvesting chlorophyll antenna size in *Chlorella vulgaris* improves biomass productivity. J Appl Phycol 28: 3193-3202.

Shin W., Lee B., Kang N. K., Kim Y., Jeong W., Kwon J., Jeong B., Chang Y. K. (2017) Complementation of a mutation in *CpSRP43* causing partial truncation of light-harvesting chlorophyll antenna in *Chlorella vulgaris*. Sci Rep 7:17929.

Smith S. R., Abbriano R. M., Hildebrand M. (2012) Comparative analysis of diatom genomes reveals substantial differences in the organization of carbon partitioning pathways. Algal Res 1(1): 2-16.

Sumiya N., Fujiwara T., Era A., Miyagishima S. (2016) Chloroplast division checkpoint in microalgae. PNAS 113(47): E7629-E7638.

Trentacoste E. M., Shrestha R. P., Smith S. R., Gle C., Hartmann A. C., Hildebrand M., Gerwick W. H. (2013) Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth. PNAS 110(49): 19748-19753.

Wilhelm C., Buchel C., Fisahn J., Goss R., Jakob T., LaRoche J., Lavaud J., Lohr M., Riebesell U., Stehfest K., Valentin K., Kroth P. G. (2006) The regulation and nutrient assimilation in diatoms is significantly different from green algae. Protist 157(2): 91-124.

CONCLUSIONS

Microalgae present a promising solution to the increasing need for sustainable production of food, fuel, and chemical products [Mata et al. 2010]. The objective of this dissertation was to contribute information towards the goal of improving microalgal productivity and therefore making microalgal production economically competitive. One issue that was addressed is the disconnect often observed between the performance of microalgal strains in the laboratory versus production conditions (Chapter 1). The other issue was the suboptimal light utilization efficiency in dense microalgal cultures, which is currently a major hurdle to the economic viability of large-scale microalgal production (Chapters 2 and 3).

Promising microalgal strains that have been developed or screened in the laboratory often fail to deliver when tested in production conditions [e. g. de Mooij et al. 2015]. This has to do with differences in numerous variables, including light and temperature. In the laboratory, light and temperature are often constant, or the light is switched on and off abruptly. To take advantage of sunlight, most commercial microalgal cultivation happens outdoors, where light and temperature follow sinusoidal diel patterns. Changes in temperature and light have been found to affect microalgal growth and productivity [Bonnefond et al. 2016, Edmundson and Huesemann 2015, Ogbonna and Tanaka 1996, Orefice et al. 2016, Yang et al. 2016], and it is thus important to understand diel metabolic and physiological changes microalgal cultures experience when cultivated outdoors in order to optimize their performance. Chapter 1 followed cell cycle progression and changes in optical density (proxy for biomass), triacylglycerol (TAG, neutral lipid of interest for biofuel production), and photopigments in the production candidate diatom *Cyclotella cryptica* subject to diel sinusoidal changes in light and temperature, mimicking outdoor conditions. A major finding was that hypothetical product yields differ substantially throughout the day. This relates to the timing of light energy availability, as well as various physiological and metabolic processes such as synchronous cell cycle progression typical of microalgal cultures grown on a

light/dark cycle. Generally, TAG and biomass yields would be greatest towards the end of the light period, regardless of the species and conditions used. This is because at that time, carbon fixation for the day would be complete, but night-time energy-requiring processes would not have yet started. In our system, harvesting in the evening rather than in the morning would have increased TAG yield 4-6-fold, biomass yield 2-3-fold, and photopigment yield 1.5-3-fold. Taxonomic differences that affect the timing of various processes, as well as variables that need to be optimized for maximal yields in a microalgal production system were discussed.

Additionally, photopigment dynamics were recorded with unprecedented resolution, providing insight into how their cellular abundance responds to cell cycle progression (and therefore chloroplast division) and irradiance changes. Notably, the major diatom photoprotective pigment diadinoxanthin (Ddx), together with its de-epoxidized form diatoxanthin (Dtx), followed the sinusoidal changes in light intensity closely, while the abundance of other photopigments was affected by cell cycle progression. These results indicate that that the abundance of Ddx+Dtx is regulated separately from the main accessory light-harvesting photopigment fucoxanthin (Fx). This is intriguing, because Ddx/Dtx and Fx are end products of a branched carotenoid biosynthesis pathway that is still poorly understood in diatoms [Bertrand 2010, Coesel et al. 2008].

Chapters 2 and 3 focused on improving light utilization efficiency, and therefore productivity, in diatom cultures. Dense microalgal cultures suffer from uneven light distribution because microalgae evolved very efficient light harvesting and dissipation of excess light energy. Cells closest to the light source absorb too much light and dissipate what they do not use, and cells deeper into the culture are shaded [de Mooij et al. 2015]. As a solution, reducing the light-harvesting or dissipating capacity of microalgal cells has been explored, with some success, mostly in chlorophytes. Diatoms differ in light-harvesting and dissipation strategies from chlorophytes and are highly productive [Hildebrand et al. 2012, Wilhelm et al. 2006]. Thus, we aimed to explore

improving light utilization efficiency in diatom cultures as a means of further increasing productivity. Because the main diatom accessory light-harvesting pigment Fx and the pigment pool responsible for photoprotection and dissipation of excess light energy (Ddx/Dtx) are products of understudied diatom carotenoid biosynthesis [Bertrand 2010, Coesel et al. 2008], Chapter 2 focused on elucidating the pathway in order to identify genetic manipulation targets. The model diatom *Thalassiosira pseudonana* was chosen based on the availability of genomic, transcriptomic, and physiological data, as well as a suite of genetic manipulation tools. A major finding was the identification of a novel violaxanthin de-epoxidase-like enzyme (VDL2, Thaps3_11707), that participates in Fx biosynthesis. Heretofore, no enzymes responsible for Fx biosynthesis have been identified. Furthermore, this is the first documented function for a VDL. VDLs have been hypothesized to participate in photoprotective xanthophyll cycling wherein pigments are interconverted by epoxidation and de-epoxidation based on their similarity to the enzyme violaxanthin de-epoxidase that catalyzes the de-epoxidation reactions [Bertrand 2010]. Our findings do not support that hypothesis and demonstrate a separate role for a VDL. Additionally, another candidate that may also participate in Fx biosynthesis was identified (Thaps3_10233).

Another major finding was that reducing Fx results in an overall reduction in photopigments including chlorophylls, which are products of a separate biosynthetic pathway. Ratios of photopigments to each other were not changed. We hypothesize that Fx plays a role in the assembly and stabilization of diatom light-harvesting complexes, as other carotenoids have been found to do in other organisms [Moskalenko and Karapetyan 1996, Santabarbara et al. 2013]. One outcome of this finding is it appears to not be possible to selectively reduce Fx abundance without affecting the content of other photopigments as a possible strategy to improve light utilization efficiency. Three separate lines silencing VDL1, VDL2, and both copies of LUT1-like (LTL) simultaneously with antisense resulted in the overall reduced photopigmentation phenotype. Thus,

another outcome of this finding is that it appears that knocking down the expression of carotenoid biosynthesis genes will, with few exceptions discussed in Chapter 2, not be helpful in elucidating the function of their products, unless it is possible to perform a chemical rescue. Finally, this finding implicates LTLs in diatom carotenoid biosynthesis. They have been hypothesized to replace β-carotene hydroxylase (BCH) in diatoms [Bertrand 2010], but this has not been previously assessed experimentally. While the results do not conclusively assign the LTLs to a specific step of diatom carotenoid biosynthesis, they do indicate that they are part of the pathway.

Chapter 2 also identified several evolutionary curiosities. One of those is finding that VDL2 has a domain with only very limited similarity based on sequence or predicted structure to currently known proteins. It will be interesting to characterize it, as it may catalyze previously unobserved chemistry. Another relates to the partial BCH sequence found in *T. pseudonana* and no other currently available diatom genomes. We found that it had gained a C-terminal domain with limited similarity to uncharacterized proteins in diverse organisms, and lost chloroplast targeting. The C-terminal domain is also unique to *T. pseudonana* among available sequenced diatom genomes. Why *T. pseudonana* has retained a partial BCH sequence is unknown but studying this could bring additional insight into diatom evolution. Additionally, it would be interesting to learn what its new function is, and what the novel C-terminal domain does. Finally, it appears that a portion of a chromosome containing a phytoene desaturase gene had been duplicated by insertion into a different chromosome. The duplicated region contains at least two other genes, and it is not possible given the currently available genomic information to determine the full span of the duplicated region. Nevertheless, this duplication was retained in the population, possibly conferring an evolutionary advantage.

Finally, a model for how diatom carotenoid biosynthesis may be differentially regulated in response to chloroplast division and irradiance increase was developed, based on transcriptomic

and physiological data. Not all hypotheses generated in Chapter 2 were tested, but they lay a foundation for further investigation of the still enigmatic pathway. Orthologues of all the enzymes identified in this chapter were found in the five other currently available diatom genomes. Thus, our findings are relevant to diatom carotenoid biosynthesis in general, beyond our model organism.

In Chapter 3, we took advantage of two transgenic *T. pseudonana* lines generated in Chapter 2 to evaluate the impact of altered photosynthetic pigmentation on growth and productivity. In one line, the novel VDL2 was overexpressed (OE). A reduction in the photoprotective Ddx/Dtx, which are involved in light energy dissipation, resulted, along with a stoichiometric increase in the abundance of the light-harvesting Fx. This proved to be a promising strategy for improving diatom productivity, resulting in a substantial increase in TAG and protein accumulation during exponential growth compared to wild type (WT) (up to 3.4-fold and 2-fold, respectively). This negatively correlated with light energy dissipation through non-photochemical quenching (NPQ), most of which relies of Ddx+Dtx content. It will be important to evaluate VDL2 OE performance in production conditions, utilizing the concepts discussed in Chapter 1.

VDL2 OE grew up to 7% slower than WT, which may be due to the presence of additional misplaced Fx. Because Ddx+Dtx abundance appears to be regulated separately from Fx, as discussed in Chapters 1 and 2, it may be possible to selectively reduce it. The model for differential diatom carotenoid biosynthesis regulation developed in Chapter 2 suggests that this may be accomplished by knocking down phytoene synthase 2 (PSY2). If PSY2 is indeed responsible for activating diatom carotenoid biosynthesis in response in increased irradiance as hypothesized, knocking it down should also downregulate the extent to which Ddx+Dtx accumulate in response to sinusoidal light as observed in Chapter 1, thus increasing productivity throughout the photoperiod in cultures grown outdoors. Another option to explore is targeting proteins involved in NPQ, such as LHCX3 [Hao et al. 2018].

The other transgenic line evaluated in Chapter 3 was created by simultaneously knocking down (KD) both copies of *T. pseudonana* LTL with antisense. The LTL KD lines had an overall reduction of photosynthetic pigmentation without a change in pigment ratios. This turned out to be not as conducive to improving diatom productivity as the VDL2 OE phenotype. The upside of LTL KD was that growth was comparable to WT, and there was an increase in TAG content of up to 40%, substantially less than in VDL2 OE. The downside was a reduction in protein content and apparent sustained cellular stress. Protein can be co-harvested along with TAG as a high-value co-product, and VDL2 OE is poised to provide both in substantially higher quantities than WT. LTL KD, on the other hand, would provide a comparably modest increase in TAG yield with a concomitant decrease in protein. Additionally, the stress observed in LTL KD may result in reduced fitness in production conditions.

To summarize, Chapter 2 delved into the diatom carotenoid biosynthesis pathway, identifying potential targets for altering cellular photopigment content in hopes of improving light utilization efficiency, and therefore productivity, in dense cultures. Chapter 2 has substantially advanced the understanding of diatom carotenoid biosynthesis and identified numerous next steps that need to be undertaken to further elucidate the biosynthetic pathway. Chapter 3 investigated the effectiveness of two photopigment manipulation strategies identified in Chapter 2 for increasing productivity in diatom cultures. One of those strategies was found to be promising. The utility of that strategy for improving performance in production conditions can be evaluated using the concepts presented in Chapter 1. Taken together, this work advances our understanding of diatom metabolism and provides strategies for a potentially substantial improvement in product yields, one based on improved light utilization efficiency in dense cultures, the other based on understanding and taking advantage of diel changes in metabolic and physiological processes in cultures grown outdoors.

REFERENCES FOR THE CONCLUSIONS

Bertrand M. (2010) Carotenoid biosynthesis in diatoms. Photosynth Res 106: 89-102.

Bonnefond H., Moelants N., Talec A., Bernard O., Sciandra A. (2016) Concomittant effects of light and temperature diel variations on the growth rate and lipid production of *Dunaliella salina*. Algal Res. 14: 72-78.

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(8): e2896.

de Mooij T., Janssen M., Cerezo-Chinarro O., Mussgnug J. H., Kruse O., Ballottari M., Bassi R., Bujaldon S., Wollman F., Wijffels R. H. (2015) Antenna size reduction as a strategy to increase biomass productivity: a great potential not yet realized. J Appl Phycol 27: 1063-1077.

Edmundson S. J., Huesemann M. H. (2015) The dark side of algae cultivation: characterizing night biomass loss in three photosynthetic algae, *Chlorella sorokiniana*, *Nannochloropsis salina* and *Picochlorum* sp. Algal Res. 12: 470-476.

Hao T., Jiang T., Dong H., Ou L., He X., Yang Y. (2018) Light-harvesting protein Lhcx3 is essential for high light acclimation of *Phaeodactylum tricornutum.* AMB Expr 8: 174.

Hildebrand M., Davis A. K., Smith S. R., Traller J. C., Abbriano R. (2012) The place of diatoms in the biofuels industry. Biofuels 3(2):221-240.

Mata T. M., Martins A. A., Caetano N. S. (2010) Microalgae for biodiesel production and other applications: a review. Renew Sust Energ Rev 14: 217-232.

Moskalenko A. A., Karapetyan N. V. (1996) Structural role of carotenoids in photosynthetic membranes. Z Naturforsch 51c: 763-771.

Ogbonna J. C., Tanaka H. (1996) Night biomass loss and changes in biochemical composition of cells during light/dark cyclic culture of *Chlorella pyrenoidosa*. J Ferment Bioeng. 82(6): 558-564.

Orefice I., Chandrasekaran R., Smerilli A., Corato F., Caruso T., Casillo A., Corsaro M. M., Piaz F. D., Ruban A. V., Brunet C. (2016) Light-induced changes in the photosynthetic physiology and biochemistry in the diatom *Skeletonema marinoi*. Algal Res. 17: 1-13.

Santabarbara S., Casazza A. P., Ali K., Economou C. K., Wannathong T., Zito F., Redding K. E., Rappaport F., Purton S. (2013) The requirement for carotenoids in the assembly and function of photosynthetic complexes in *Chlamydomonas reinhardtii*. Plant Physiol 161: 535-546.

Wilhelm C., Buchel C., Fisahn J., Goss R., Jakob T., LaRoche J., Lavaud J., Lohr M., Riebesell U., Stehfest K., Valentin K., Kroth P. G. (2006) The regulation and nutrient assimilation in diatoms is significantly different from green algae. Protist 157(2): 91-124.

Yang W., Zou S., He M., Fei C., Luo W., Zheng S., Chen B., Wang C. (2016) Growth and lipid accumulation in three *Chlorella* strains from different regions in response to diurnal temperature fluctuations. Bioresour Technol. 202: 15-24.