# ABSTRACT


| | |
|---|---|
| Title of Dissertation: | INSIGHTS INTO DINOFLAGELLATE NATURAL PRODUCT SYNTHESIS VIA CATALYTIC DOMAIN INTERACTIONS |
| | Ernest Patrick Williams, Doctor of Philosophy, 2022 |
| Dissertation directed by: | Professor Allen Richard Place, Institute of Marine and Environmental Technologies, University of Maryland Center for Environmental Science |



Dinoflagellates are protists that can be split into two evolutionary groups, the parasitic syndinians and the largely photosynthetic "core" dinoflagellates. They represent a major portion of aquatic biomass which means that they are responsible for large portions of carbon that are both fixed and released. Other than biomass, the fixed carbon can be made into natural products such as polyunsaturated fatty acids that support the biota of many ecosystems or toxins that are harmful to aquatic life and humans. DNA and RNA analyses have been used to discover the putative genes that may make these compounds, but their non-colinear arrangement in the genome is very different from model organisms and their gene copy number is very high, making it nearly impossible to determine the exact biosynthetic pathways. The goal of my studies was to develop methods to differentiate biosynthetic pathways such as lipid and toxin synthesis by comparing the ability of domains to interact with each other with the assumption that domains that preferentially interact are more likely to participate in the same pathway. Initially, a survey

was performed on available dinoflagellate transcriptomes to enumerate domains potentially involved in natural product synthesis and bin them based on sequence similarity to identify genes that could be used in biochemical assays. An interesting integration of analogous genes involved in lipid synthesis with those involved in natural product synthesis was observed as well as trends in domain expansion and contraction during core dinoflagellate evolution. Ultimately, the domain that scaffolds natural product synthesis, the thiolation domain, was chosen for further study because it exhibited two clear functional bins and is acted on directly by another enzyme, a phosphopantetheinyl transferase (PPTase). The PPTase activates the thiolation domain by transferring the phosphopantetheinate group from Coenzyme A to the thiolation domain, creating a free thiol group upon which the natural products are synthesized. These PPTases were then enumerated in dinoflagellates and characterized by looking for sequence motifs and observing expression patterns over a diel cycle as well as during growth in the model species *Amphidinium carterae*, a basal toxic dinoflagellate. *Amphidinium carterae* appears to have three PPTases, two of which (PPTase 1 and 2) are very similar, except that PPTase 2 does not appear to have a stop codon and has never been observed as a full-size protein. The remaining two PPTases (PPTase 1 and 3) had alternating expression patterns that did not appear to directly correlate to the acyl carrier protein, the thiolation domain required specifically for lipid biosynthesis. This carrier protein, like other enzymes for natural product synthesis in dinoflagellates, had a chloroplast targeting sequence while the three PPTases did not. To investigate the ability of these three PPTases to activate various thiolation domains, a total of 8 domains from *A. carterae* were substituted into the blue pigment synthesizing gene BpsA from *Streptomyces lavendulae*. These recombinant constructs were used for coexpression in *E. coli* as well as *in vitro* to reduce as many artifacts as possible and assess the interactions of each PPTase with the thiolation domains.

Some of the recombinant BpsA genes were able to make blue dye with all three PPTases, while others never made blue dye both in *E. coli* as well as in vitro. In vitro quantification of free thiol added by the PPTase showed that all the thiolation domains, as well as the acyl carrier protein could be phosphopantetheinated by all the PPTases. This generalist substrate recognition, along with the alternating expression patterns and lack of chloroplast signaling peptide, indicate that the two active PPTases are performing the same function on all available thiolation domains, probably before export to the chloroplast. This lack of pathway segregation by PPTases is a completely novel way of synthesizing natural products compared to bacteria and fungi, likely due to the acquisition of both photosynthesis and natural product/lipid biosynthesis during dinoflagellate evolution that was not present in the common ancestor. Additionally, the techniques to identify genes of interest and perform biochemical characterization developed here are useful for future experiments annotating the function of dinoflagellate genes.

# INSIGHTS INTO DINOFLAGELLATE NATURAL PRODUCT SYNTHESIS VIA CATALYTIC DOMAIN INTERACTIONS

by

Ernest Patrick Williams

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2022

Advisory Committee:
Professor Allen Richard Place, Chair
Tsvetan Radislavov Bachvaroff
Yantao Li
Michael Gonsior
Jum Sook Chung

# Preface

Dinoflagellates are incredibly interesting to study but also incredibly frustrating because of their unique biology. Every experiment seems to create more questions than it answers. Many of these studies focus on how to work on dinoflagellates rather than the results themselves. I've learned to step away from canonical thinking and let the biology of the dinoflagellates speak for themselves. They are not similar to model organisms, nor are they primitive or simple. They have their own biology that's been evolving for hundreds of millions of years. A frequent joke is that if it's been shown in yeast, mouse, and human then it must occur in all of eukaryotic life. Researchers who say that have never studied dinoflagellates.

# Dedication

I'd like to thank the Ackerley lab from the University of Victoria Wellington for their donation of the BpsA gene. I'd also like to thank my family for their support during this process.

# Acknowledgements

This work was funded by NSF, NIEHS: Oceans, Great Lakes and Human Health RO1, Maryland Industrial Partnerships. The Bailey Wildlife Foundation, Maryland Department of Agriculture, The IMET Angel Investors, USDA Agriculture and Food Research Initiative Competitive Grants Program, MEES graduate student funding, and IMET Innovation funds. This work took place at the Institute for Environmental Sciences, a part of the University of Maryland Centers for Environmental Science. Support was provided by Luke Freeney during his internship at IMET. Sequencing support was provided by Sabeena Nazar as part of the Bio-Analytical Services Laboratory at IMET.

# Table of Contents

This Table of Contents is automatically generated by MS Word, linked to the Heading formats used within the Chapter text.

# List of Tables

Table 3-1: Primers and oligonucleotides used in the construction and insertion of dinoflagellate thiolation domains into the BpsA reporter.

# List of Figures

**Introduction**



**Chapter 1**

# List of Abbreviations

PPTase: phosphopantetheinyl transferase

ACP: acyl carrier protein

PKS: polyketide synthesis

NRPS: non-ribosomal peptide synthesis

# Introduction

## Overarching goals

As major primary producers in the world's oceans and one of the few producers of certain polyunsaturated fats, dinoflagellates are vitally important to the global ecosystem. They are simultaneously detrimental to many ecological consumers, including humans, via the production of toxins that accumulate at various trophic levels. Despite intense efforts to understand their ecological impact, we know very little about the biochemistry of how these compounds are made. Many good-faith predictions have been made about which genes participate in their biosynthesis, but they are based on models that are separated from dinoflagellates by a huge evolutionary distance because of the lack of information on protists in general.

Also, the redundancy of chemistries used in natural product synthesis and the potential for iterative processes make predictions especially challenging. Thus, a bottoms-up approach is necessary to root further predictions in evidence that is based on dinoflagellate evolutionary biology.

The thesis aims to advance understanding of how dinoflagellates synthesize natural products with a specific focus on separating toxin and other metabolic synthetic pathways from each other and from the essential and chemically identical process of synthesizing lipids. The overall rationale is to use dinoflagellate-specific datasets to designate genes for biochemical characterization. This approach differs from previous characterizations of dinoflagellate genes in that as few assumptions as possible were made based on model organisms. This reduces bias and allows us to acknowledge that dinoflagellates are strange organisms that break the rules at every turn. Instead, enumeration and binning of all dinoflagellate synthetic domains will allow us to identify tractable candidates that can be examined on a case-by-case basis. Biochemical analyses rely on heterologous expression of dinoflagellate proteins in *E. coli* which allows for protein interactions to occur in a prescribed setting to reduce as many artifacts as possible. Determining the specificity of these interactions is the basis for the underlying assumption that synthetic pathways are segregated molecularly in dinoflagellates. The following studies will attempt to establish whether or not this is true.

## Dinoflagellate taxonomy and evolution

Dinoflagellates are unicellular protists that can be split into two major evolutionary lineages, the early branching heterotrophic and empirically parasitic syndinians and the largely photosynthetic "core dinoflagellates" (Bachvaroff et al., 2014; Hoppenrath & Leander, 2010; Janouškovec et al., 2017; Taylor, Hoppenrath, & Saldarriaga, 2008) (Figure I-1). Syndinian dinoflagellates have not

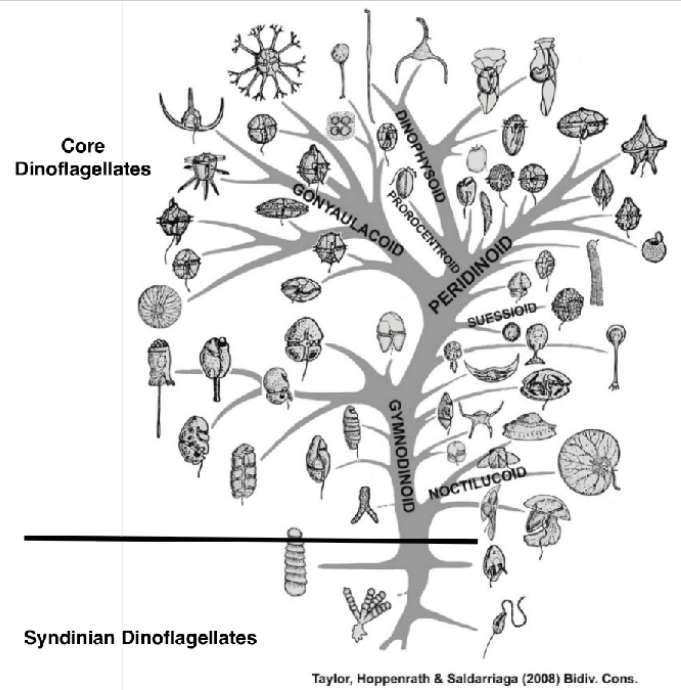
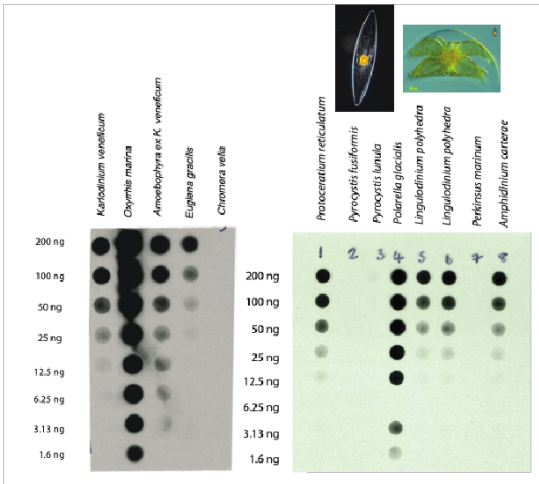



**Figure I-1: Illustration of dinoflagellate relationships**
Shown is an illustrated version of the relationships of various dinoflagellate lineages with representative species depicted from Taylor *et al.*, 2008. The black line separates the syndinian dinoflagellates on the bottom and the core dinoflagellates on the top.

Shown is an illustrated version of the relationships of various dinoflagellate lineages with representative species depicted by Taylor *et al* 2008. The black line separates the syndinian dinoflagellates on the bottom and the core dinoflagellates on the top.

been widely studied with only a few taxonomically described species that infect fish eggs, ciliates, and even other dinoflagellates. (Bachvaroff, Kim, Guillou, Delwiche, & Coats, 2012; Coats, Bachvaroff, & Delwiche, 2012; Harada, Ohtsuka, & Horiguchi, 2007; Miller, Delwiche, & Coats, 2012; Skovgaard, Meneses, & Angélico, 2009). They are ubiquitous in the upper ocean and represent a huge fraction of marine microbial diversity (de Vargas et al., 2015). Although all observed syndinians are not photosynthetic, their evolutionary history is under some debate. There is some evidence for plastid genes in the *Hematodinium* sp. genome leading to the hypothesis that the common ancestor is photosynthetic, and this organelle has been lost in all extant syndinian lineages (Gornik et al., 2015). Other evidence includes the relic plastids in the closely related apicomplexans (Gajadhar et al., 1991) as well as extant photosynthetic basal species (Moore et al., 2008; Oborník et al., 2012). The most basal members of syndinians, the marine alveolate group I, are not evidently photosynthetic, making a common photosynthetic ancestor unlikely.

Horizontal gene transfer that seems to be a hallmark of dinoflagellate evolution (Wisecaver, Brosnahan, & Hackett, 2013), is a more likely explanation of foreign DNA sequences in *Hematodinium* sp., starting with the acquisition of a nucleoprotein likely of viral origin termed the dinoflagellate viral nucleoprotein (DVNP) that coincides with atypical chromatin structure in marine alveolate groups II and IV including *Hematodinium* sp. (Gornik et al., 2012; Strassert et al., 2018). The core dinoflagellates have various biological features separating them from syndinian groups as well as changes specific to certain core dinoflagellate lineages (Janouškovec et al., 2017). The most striking is the "dinokaryon", a term applied to the unique physical and biochemical properties of core dinoflagellate chromosomes (Fukuda & Suzaki, 2015; Gornik, Hu, Lassadi, & Waller, 2019). One feature is reducing histones that package the genome (Marinov & Lynch, 2015), which are largely replaced by a major basic nuclear protein (Kato et al., 1997). This is coincident with a high concentration of divalent cations in the chromatin (Levi-Setti, Gavrilov, & Rizzo, 2008; Moreno Díaz de la Espina, Alverca, Cuadrado, & Franca, 2005) and a replacement of a large portion of thymidine residues with 5′ hydroxymethyl-uracil (Rae, 1973; Williams & Place, 2014) as well as the glucosylated form "Base J" that controls transcription termination in trypanosomes (Figure I-2). Physically these changes to the chromatin




| | %5hmU | Relative base J |
|---|---|---|
| K. veneficum | 52.7 | + |
| O. marina | 0 | +++ |
| Amoebophyra ex K.v. | 0 | ++ |
| Euglena gracilis | 1.0 | + |
| Chromera velia | 0 | - |
| Protoceratium | 64.3 | + |
| Pyrocystis lunula | 24.4 | - |
| Polarella glacialis | 44.5 | ++ |
| L. polyedra | 26.7 | + |

**Figure I-2: 5-hydroxymethyl uracil and base J in dinoflagelates**
The right panel shows the amounts of 5-hydroxymethyl uracil in several dinoflagellate and alveolate outgroups as a percentage of total thymidine derived bases. The left panel is an immunoblot probing the glucosylated form of 5-hydroxymethyl uracil called "base J" in five-fold dilutions with two example *Pyrocystis* species shown above.

composition result in a unique morphology where the chromosomes are condensed with a left-handed screw and steep pitch not observed with the canonical nucleosomal structure (Costas & Goyanes, 2005) to the point where protein binding is altered (Potapov et al., 2018) and birefringence has been observed (Chow, Yan, Bennett, & Wong, 2010; Livolant, 1978) (Figure I-3). Thus, the dinokaryon chromosome is

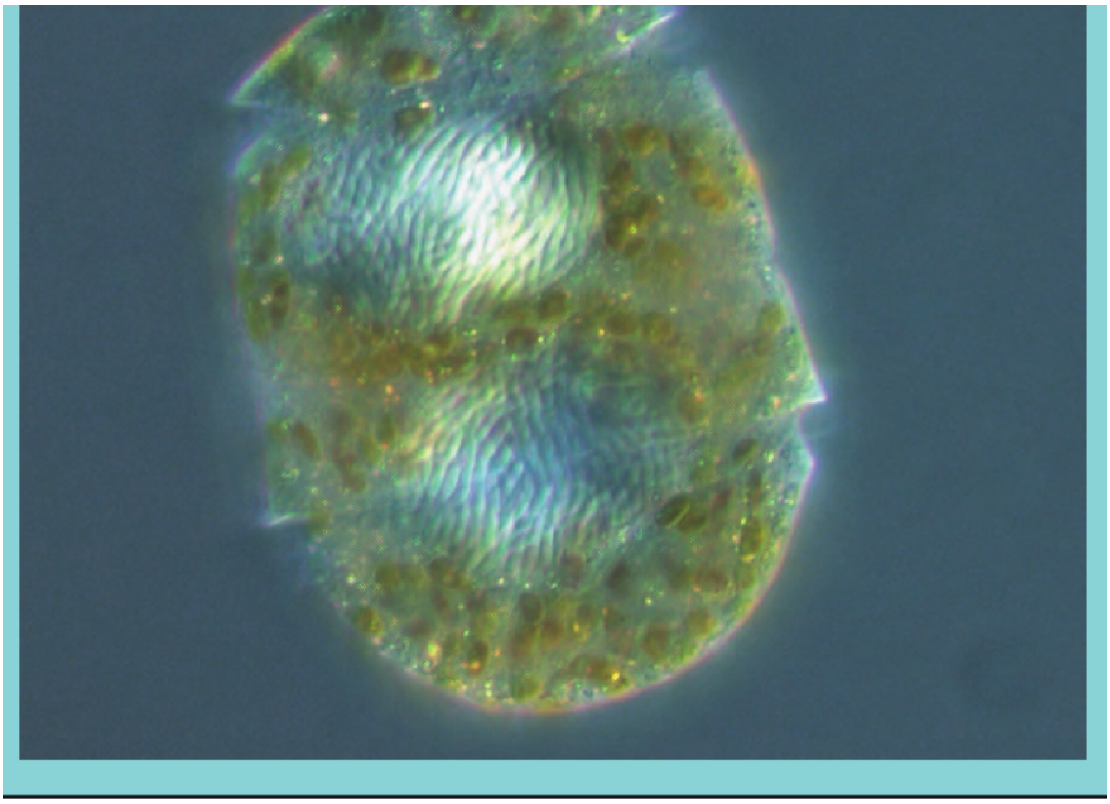


**Figure I-3: Dinoflagellate chromosome birefringence**
A dinoflagellate *Polykrikos hartmanii* cell is shown highlighting the chromosomes labeled the "dinokaryon". The arched fibril structure can be seen for individual chromosomes.

frequently described as "liquid crystalline" and was originally compared to the polytene chromosomes of *Drosophila*, where stacks of identical sequences also give a crystalline appearance (Zykova, Levitsky, Belyaeva, & Zhimulev, 2018).

Apart from the nuclear features that are synapomorphic to the "core dinoflagellates" is the observation of multiple chloroplast gains and losses not seen in other alveolates and beginning with the likely engulfment of a haptophyte, another photosynthetic eukaryote resulting in four chloroplast membranes (Janouskovec, Horák, Oborník, Lukes, & Keeling, 2010; Sato, 2020; Tengs et al., 2000; Yoon, Hackett, & Bhattacharya, 2002). Photosynthetic dinoflagellates were originally characterized as having a unique pigment named peridinin (Carbonera, Di Valentin, Spezia, & Mezzetti, 2014) that complexes with chlorophyll α in a pigment/protein complex that can be traced back to a haptophyte origin of the most basally observed chloroplast (Bachvaroff, Sanchez Puerta, & Delwiche, 2005; Yoon et al., 2002). Even this has been in some dispute, with certain sequences giving different phylogenies than others (Bachvaroff, Sanchez Puerta, & Delwiche, 2005), providing an example of the understatement that fundamental genetic changes occur following an endosymbiotic event (Dorrell & Howe, 2015). Non-photosynthetic species, including the genera *Oxyrrhis* and *Noctiluca,* have retained specific plastid-derived metabolic pathways, indicating the acquisition of a chloroplast was a defining moment in

dinoflagellate evolution (Janouškovec et al., 2017). Subsequent chloroplast acquisitions retained over large spans of evolutionary time from the engulfing of other photosynthetic protists (Yamada, Sym, & Horiguchi, 2017; Yamada et al., 2019) are a demonstration of the incredible genomic plasticity dinoflagellates possess given that almost every other chloroplast acquisition is either extremely rare or transient (Hehenberger, Gast, & Keeling, 2019; Maselli, Anestis, Klemm, Hansen, & John, 2021; Pelletreau et al., 2011). It is unlikely a coincidence that genomic plasticity and the relative ease of chloroplast acquisition are common to the core dinoflagellates and should be considered when investigating gene function and evolution.

The taxonomy of syndinian dinoflagellates is almost entirely based on molecular data since so few species have been cultured, and the bulk of the information is based on environmental DNA surveys (de Vargas et al., 2015) and a limited set of genomes (Bachvaroff, 2019; John et al., 2019). In a general sense, syndinians are sometimes grouped based on the host organisms that they parasitize (Coats et al., 2012; Miller et al., 2012; Probert et al., 2014; Skovgaard et al., 2009), but this too is based on an extremely limited set of described organisms. Core dinoflagellates on the other hand have frequently been characterized and recharacterized based on their thecal plate arrangements (Dodge, 1995; Gómez & Artigas, 2019; Iwataki, 2008) for "armored" dinoflagellates, and cellular ultrastructure for "naked" dinoflagellates (Blanco & Chapman, 1987). This is not surprising since in-depth investigations into dinoflagellates and their biology coincided with the use of electron microscopy (DODGE & CRAWFORD, 1970). These morphological characters have been instrumental in informing sequence-based phylogenies resulting in fairly robust descriptions of the lineages within the core dinoflagellates (Bachvaroff et al., 2014; Janouškovec et al., 2017).

When investigating the biology of the core dinoflagellates, it's easy to forget about the syndinian lineages because they are so poorly studied and seemingly so different. The complexity and strangeness of the core dinoflagellates, however, make their outgroup even more important to place observations into a larger context to help make sense of their unique biology. The incredible differences in nuclear physiology and endosymbiotic history between the core and syndinian dinoflagellates helps to understand how and why many biological differences may arise and prevent lumping core dinoflagellate biology together with other photosynthetic organisms such as plants or even other protists such as diatoms that span a huge evolutionary space and distance. Yes, dinoflagellates are strange, but they didn't come from outer space (at least not likely) and their evolutionary history is an important tool for understanding their biology.

### Dinoflagellate replication and circadian rhythms

Both syndinian and core dinoflagellates have an identified life stage as a free-living "dinospore" with two flagella that provide motility (DODGE & CRAWFORD, 1970; Miller et al., 2012). Syndinians also exist as an intracellular parasite of their host, with a variety of morphologies and strategies exhibited between species (Coats

et al., 2012; Harada et al., 2007; Skovgaard et al., 2009; Miller et al., 2012) (Figure I-4). The parasitic life stage is regarded as distinct from the free-living stage and is

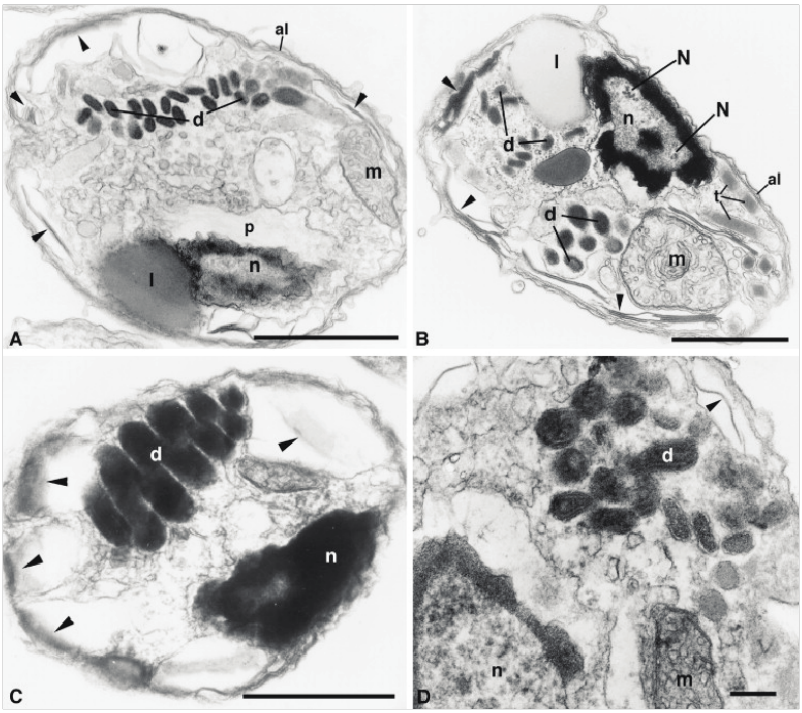



**Figure I-4: Stages of *Akashiwo sanguinea* infection by *Amoebophrya sp. ex A. sanguinea***

Micrographs are shown of an Akashiwo sanguinea cell infected by the syndinian dinoflagellate Amoebophrya sp from {Miller et al., 2012, #282455}. The "vermiform" stage can be seen in panel "C" where multiple cells are connected following replication but prior to eruption from the host.

when syndinians replicate following the consumption of host resources. Core dinoflagellates, on the other hand, have a mix of sexual and resting stages in addition to their free-living stage, making for a potentially complex life cycle that is sometimes species-specific (Coats, Tyler, & Anderson, 1984; Lee, Chiang, & Tsai, 2021; Litaker et al., 2002; Warns, Hense, & Kremp, 2013). *Pyrocystis* is an example of a core dinoflagellate that exists primarily in the cyst stage, with flagellate stages observed as an intermediate (Pincemin & Gayol, 1978), while other species occur primarily as free-living flagellates.

For syndinians, replication primarily results from infection and is not relegated to certain times of the day. They are usually found in the photic zone, however, and autofluorescent, indicating that they may be able to sense light. Thus, their circadian rhythms are likely tied predominantly to their hosts from which they derive their nutrients (Yih & Coats, 2000). Otherwise, they are quite normal, having a classic replicative pattern with multiple growth phases followed by a synthesis phase and mitosis. Synthesis and mitosis frequently occur within the host with several morphologies of how and when the replicating cells dissociate (Coats & Park, 2002).

Often the replicated cells will remain attached to one another and come apart to become free-living dinospores upon eruption from their host.

Core dinoflagellates also exhibit a canonical cell cycle, with haploid cells dividing by mitosis following a synthesis phase of nucleic acids (Bhaud et al., 2000). Mitosis, on the other hand, is quite strange in core dinoflagellates. The chromatin remains condensed, and the nuclear envelope remains intact without spindle fiber formation (Tippit & Pickett-Heaps, 1976). Instead, microtubules extend through an extensive network of tunnels in the membrane (Gavelis et al., 2019) and chromosomes are separated within the nucleus without an apparent polar region, followed by invagination of the nuclear membrane and separation to the daughter cells (Oakley & Dodge, 1976). There are some exceptions, such as in *Oxyrrhis marina,* where polar separation is apparent (Gao & Li, 1986). It is unclear if these particular methods of mitotic division are related to the unique chromosomal attributes of the core dinoflagellates or simply coincident with the advent of the dinokaryon.

Observations of cell cycle and division in dinoflagellates have been made possible through the synchronization of the cell cycle by modifications in the light-dark cycle (Galleron, 1976). This indicates that basic biological processes in dinoflagellates operate on a circadian cycle. This includes nutrient cycling, division, bioluminescence, and toxin production (Jia, Gao, Tong, & Anderson, 2019) and also appears to be tied to photosynthesis. This would indicate that a heavy reliance on circadian rhythms may have come about following the acquisition of a plastid. There is no strong evidence for circadian rhythmicity in syndinians, but again it must be noted that a very limited amount of study has been done on their basic biology. One of the most frequently studied examples of a circadian controlled process in core dinoflagellates is bioluminescence and its concomitant rhythms within the chloroplast (Hardeland & Nord, 1984). This bioluminescence results from luciferase controlled by the luciferin binding protein (Morse & Mittag, 2000) that occurs in an organelle called the scintillon (Desjardins & Morse, 1993). The actual activation of the scintillon is a pH- mediated process mediated by a mechanosensing proton channel (Rodriguez et al., 2017). When a scintillon-containing cell is disturbed, the mechanosensing channel allows protons to enter, dropping the pH and allowing luciferin to act and produce light. This only occurs at night and is based on the expression of the diurnally regulated luciferin binding protein. This regulation has been associated with a 3′ untranslated sequence that, when blocked, will disrupt the expression pattern of the luciferin binding protein (Lapointe & Morse, 2008). Many other genes are expressed on a circadian rhythm, but not all (Markovic, Roenneberg, & Morse, 1996), and the purpose of rhythmic expression is not entirely clear. Binding protein expression solely to day or night and not responding to an environmental condition seems sub-optimal. One theory by J. Woodland (Woody) Hastings, a long-time researcher of circadian rhythms in dinoflagellates, postulated that it is a mechanism for cycling amino acid nitrogen availability in the cell (Hastings, 2013). Nitrogen availability is also tightly linked to photosynthetic efficiency since it is required to make the reductant NADP and a reduction in amino acid nitrogen availability directly results in algal photosynthesis (Cointet et al., 2019; Zhao et al., 2017). This is an essential time to stop and remember that syndinians are

heterotrophic and that core dinoflagellates are often mixotrophic, getting nitrogen from ingestion of prey. Thus prey ingestion also impacts photosynthetic efficiency (Dagenais-Bellefeuille & Morse, 2013; Hansen, Skovgaard, & Stoecker, 2000). Locking nitrogen into proteins and releasing them regularly may be a way of protecting nitrogen reserves in the cell during lean times via circadian regulation of gene expression to allow for photosynthesis to proceed efficiently and promote survival.

### Dinoflagellate genomic arrangement and the regulation of gene expression

Another trait that is synapomorphic in the core dinoflagellates is expanded genome size. Many estimates place the number much higher than the human genome, with a size ranging from a few picograms to 280 picograms (Du et al., 2016; Hong et al., 2016; Veldhuis & Kraay, 2000; Allen, Roberts, Loeblich, & Klotz, 1975). Genome sizes appear to expand when investigating more distal lineages in core dinoflagellates with early-branching clades and the symbiotes of corals having the smallest genome (Aranda et al., 2016; LaJeunesse, Lambert, Andersen, Coffroth, & Galbraith, 2005b). Syndinians, in comparison, have very small genomes in the 100 Mb size range, and larger than many apicomplexans with genome sizes in the 10Mb range. Genomic expansion does not appear to correlate with expanded functionality but rather extreme levels of gene duplication (Hou & Lin, 2009). The genes themselves are seemingly arranged as tandem repeats (Bachvaroff & Place, 2008), although it is unclear if all copies are expressed or function equivalently.

Dinoflagellates perform transcription in the same way as other eukaryotes but seem to be lacking specific transcription factors, only possessing general transcription factors for the binding of the RNA polymerase complex (Roy & Morse, 2013). There is also no evidence of promoter sequences to initiate transcription and a general lack of DNA binding domains other than a low melting temperature region similar to a TATA box but with a non-canonical sequence (Beauchemin et al., 2012; Li & Hastings, 1998). Gene arrangement within the genome shows many many stretches of genes with the same orientation or “strandedness” as might be expected with tandemly repeated genes but also with unrelated genes (Lin et al., 2015; Shoguchi et al., 2013; Stephens et al., 2020) (Figure I-5). This has led to the notion that transcription may proceed polycistronically,

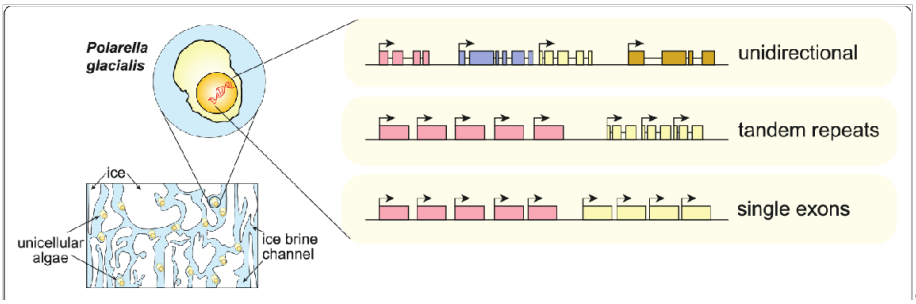



**Figure I-5: Common gene arrangements in *Polarella glacialis***
A representation of the common gene arrangements are shown from the *Polarella glacialis* genome in Stephens *et al.* 2020. Genes are frequently oriented in the same direction or "strandedness" with tandem repeats of the same genes or multiple different genes.The color of the blocks indicate a unique gene while the arrow above indicates the direction of transcription.

that multiple transcripts are processed simultaneously with a single initiation event, as demonstrated in trypanosomes (Clayton, 2019). There is not much evidence for this (Beauchemin et al., 2012), indicating that polycistronic transcription is infrequent or rapid.

In core dinoflagellates, the transcripts themselves have very long half-lives (Morey & Van Dolah, 2013), with some exceeding the time of the cell cycle, indicating that some transcripts may be passed to daughter cells during division. This is quite extraordinary when considering the central dogma of molecular biology, *i.e.,* that DNA is transcribed to RNA, which is then translated into protein. While syndinian dinoflagellates appear to follow this canon by using transcription factors as a primary means of controlling gene expression (Bachvaroff, Place, & Coats, 2009), the transcript abundance in core dinoflagellates does not correlate to protein abundance (Fagan, Morse, & Hastings, 1999; Lidie, Ryan, Barbier, & Van Dolah, 2005; Morse, Milos, Roux, & Hastings, 1989). This demonstrates a fundamental disconnect in understanding dinoflagellate biology and is a rare example of gene expression controlled post-transcriptionally. This also means that transcriptomic analyses are of limited utility and that a focus on protein expression is more useful for determining biological responses.

Micro RNAs (miRNAs) are a common means of controlling gene expression post-transcriptionally in other eukaryotes. In this process, a small RNA, usually a few dozen nucleotides, binds to a complementary 3′ sequence that then initiates degradation or can inhibit translation (O'Brien, Hayder, Zayed, & Peng, 2018). Although miRNAs are present and active in dinoflagellates as in other eukaryotes (Baumgarten et al., 2018; Gao et al., 2013; Shi et al., 2017) they represent a more nuanced approach to regulating gene expression rather than a global mechanism. MiRNAs have also been shown to not correlate with circadian expression (Dagenais-Bellefeuille, Beauchemin, & Morse, 2017) supporting circadian control, exemplified by bioluminescence, is a plastid derived process rather than host-derived (Janouškovec et al., 2017). Another common mechanism for regulating gene expression post-transcriptionally is codon bias, where the availability of specific

tRNAs is exploited to control translational efficiency (Aitken & Lorsch, 2012; Angov, Hillier, Kincaid, & Lyon, 2008; Hershberg & Petrov, 2008; Novoa & de Pouplana, 2012; Peden, 2000). The result is that certain mRNA transcripts are translated more efficiently than others, depending on the codons used in that transcript, and the abundance of specific proteins can be modulated. Again, this is a nuanced approach to controlling gene expression and is not likely to be a sole mechanism of regulation, but it is a potentially global mechanism affecting all mRNAs to varying degrees. Surprisingly (or not considering their overall biology), core dinoflagellates do not seem to have any bias at the transcript level using all codons with nearly equal frequency, except those with adenine or thymine at the third position (Williams, Place, & Bachvaroff, 2017) (Figure I-6). In

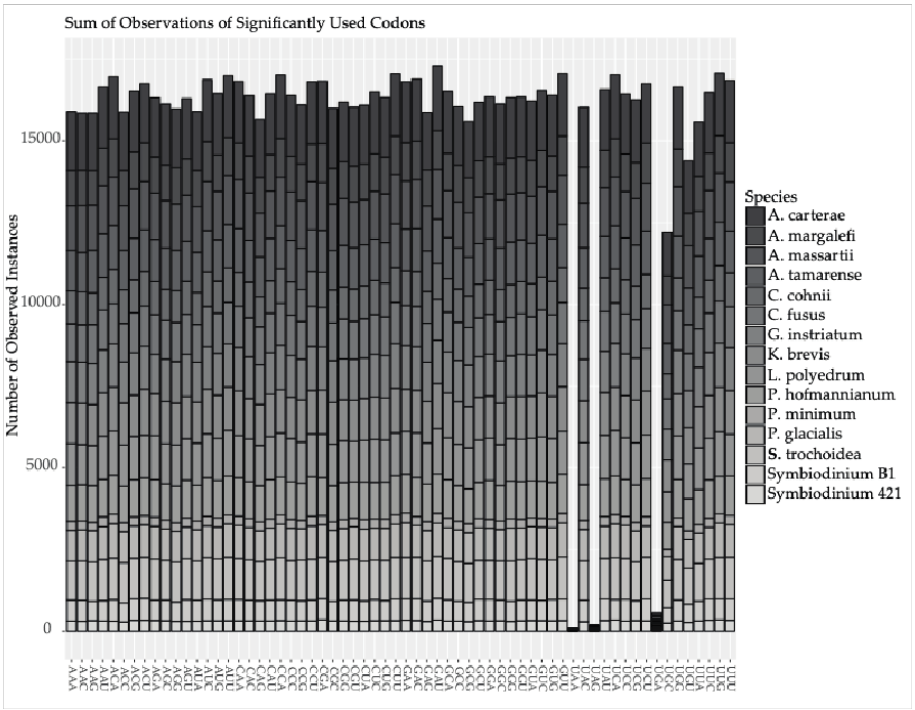



**Figure I-6: Observed codons in dinoflagellate transcriptomes**
The sums of the observations of each codon are shown for fifteen dinoflagellate transcriptomes. The X-axis describes each codon while the sum of each codon is shown on the Y-axis. Each transcriptome is colored in grayscale. Each codon is in near equal frequency in transcripts in each genome with the exceptions of the AT-rich codons UAA and UAG as well as the stop codon UGA.

some ways, this is a logical result of extreme gene duplication and subsequent point mutations, mostly at the third codon position (Bachvaroff & Place, 2008). A lack of transcriptional elements that control gene expression may also help to explain why horizontal gene transfer is observed so frequently in dinoflagellates (Wisecaver et al., 2013) as well as the integration of plastids (Keeling, 2010). Adaptation to the host's transcriptional mechanisms is necessary for the successful expression of horizontally transferred genes, and removal of those impediments may make integration more likely. Either way, a post-transcriptional approach to regulating gene

expression in core dinoflagellates should give us pause when thinking about gene function and evolution as well as the techniques we use to study them.

In terms of translation, core dinoflagellates are typical with much of the canonical machinery present (Roy, Jagus, & Morse, 2018), albeit with a higher than expected copy number (Jones, Williams, Place, Jagus, & Bachvaroff, 2015). Dinoflagellates also appear to translate mRNA using multiple ribosomes simultaneously, termed polysomes, similar to other eukaryotes (Schröder-Lorenz & Rensing, 1987). What is different in the core dinoflagellates is that the mRNAs that are usually capped on the 5′ end with a methylated base to prevent degradation and allow for translation initiation recruitment (Decroly, Ferron, Lescar, & Canard, 2011) are instead trans-spliced with a 22 base conserved nucleotide sequence that already contains the methylated cap structure (Lidie & van Dolah, 2007; Zhang et al., 2007). In trans-splicing, a portion of one stretch of nucleic acids is combined with another sequence to attach a novel sequence and is distinct from intron splicing, where a nucleic acid sequence recombines with itself to remove a region (Lasda & Blumenthal, 2011). This is termed the "spliced leader" and is a hallmark of core dinoflagellate mRNAs (Zhang, Zhuang, Gill, & Lin, 2013), being nearly ubiquitous and conserved (Zhang, Campbell, Sturm, & Lin, 2009). The spliced leader sequence is also present in both syndinian and *Perkinsus* genomes. The DNA-based sequence from which the spliced leader transcript is derived is present throughout the dinoflagellate genome (Lin et al., 2015; Shoguchi et al., 2013) and is frequently found as a retrotransposon (Song et al., 2017), likely due to its autocatalytic nature. This presents a philosophical problem, though. If all transcripts are trans-spliced, which allows translation initiation to proceed, then how can we observe long transcript half-lives? Shouldn't these transcripts be translated all the time, and if not, what's stopping it? One clue is the 3′ UTR motif associated with circadian control of gene expression (Lapointe & Morse, 2008). The result that disrupting this motif can break circadian control indicates that regulatory elements in the 3′ UTR may be inhibiting translation and are worth further investigation.

The high copy number of tandemly arranged genes in core dinoflagellates is almost certainly related to the shift towards post-transcriptional control of gene regulation since the removal of genomic elements regulating transcription allows for a more flexible genomic arrangement (and transcription may be proportional to gene copy number). Traditional approaches of measuring changes in transcript abundance during physiological treatment to determine gene function are not readily applicable to studying core dinoflagellate and instead protein quantification and biochemical assays are more likely to yield helpful results. Functional assignment based on sequence similarity is dubious if duplication has allowed for novel functionality and also points to the need for the biochemical validation of function.

### Lipids and sterols of dinoflagellate

Core dinoflagellates and diatoms have been the darling of the petroleum industry (William, 1962), because their structure is readily preserved in the fossil record (Evitt, 1961; Boere et al., 2009) and they represent a large fraction of primary

productivity based on biomass (Cloern, Foster, & Kleckner, 2014; Regaudie-de-Gioux, Lasternas, AgustÃ, & Duarte, 2014). This means that there is a huge amount of carbon flowing through the dinoflagellate lineage with core dinoflagellates as mixotrophic carbon fixers and releasers, and the parasitic syndinians also releasing carbon. Besides nucleic acids, a large fraction of cellular carbon is used to make lipids and sterols. The primary limiters of carbon fixation are the availability of carbon dioxide and photosynthetic efficiency. Eukaryotic phytoplankton will frequently modulate their biology to optimize the amount of available carbon dioxide in their chloroplast for fixation (Cecchin et al., 2021). Photosynthetic efficiency, on the other hand, is quite complex and relies on the availability of several metals, their oxidative state, available reductant, and oxidative stress (Alberts et al., 2002). The ultimate fate of the electrons produced during photosynthesis is NADP forming NADPH, while the carbon dioxide is used to carboxylate ribulose to form two 3-phosphoglycerates. Both 3PGA and NADPH participate in many processes but fundamentally are involved in the formation of acylated glycerol. Nitrogen starvation in many algae results in an immediate increase in neutral lipid synthesis (Dong et al., 2013; Li, Han, Sommerfeld, & Hu, 2011; Wang et al., 2019), presumably as a way of recycling NADP used in the reduction steps of lipid synthesis to form acyl chains that can be used in photosynthesis to reduce oxidative stress from electron accumulation.

Syndinian dinoflagellates do not possess the canonical machinery to make lipids but may make cholesterol (Leblond, Sengco, Sickman, Dahmen, & Anderson, 2006), while lipids are presumably derived intact from the hosts they infect. Core dinoflagellates, on the other hand, possess a full complement (even a surfeit) of the genes necessary for lipid synthesis (Kohli, John, Van Dolah, & Murray, 2016), albeit with some abnormal gene duplications (Williams, Bachvaroff, & Place, 2021). Specifically, lipid synthesis requires an acyl carrier protein to scaffold synthesis; a ketosynthase to attach acetate to the carrier protein from malonyl CoA created by an acetyl CoA carboxylase (Ohlrogge & Jaworski, 1997); a ketoreductase, alcohol dehydrogenase, and enoylreductase to fully reduce the carbohydrate to an acyl chain; and finally, chain length factors that are non-functional ketosynthases that inhibit further synthesis so that a thioesterase can cleave off the acyl chain. The primary free saturated fatty acid in dinoflagellates is palmitic acid (16:0) (Mansour, Volkman, Jackson, & Blackburn, 1999), which is bound to glycerol and further modified to form mono and di-galactosyldiacylglycerol (Leblond & Dahmen, 2012) that dominate the thylakoid membranes of plants (Hölzl & Dörmann, 2019). Much of the biochemistry has been worked out in *Arabidopsis thaliana* (Nilsson et al., 2015), but similar reactions appear to take place in protistan algae and are ultimately derived from cyanobacteria despite some replacements (Sato, 2020). Dinoflagellate chloroplasts are different in that they have four membranes, presumably from the previous haptophyte host that was engulfed by the ancestor of core dinoflagellates (Delwiche, 1999; Keeling, 2010) (Figure I-7), making this story somewhat more complicated.

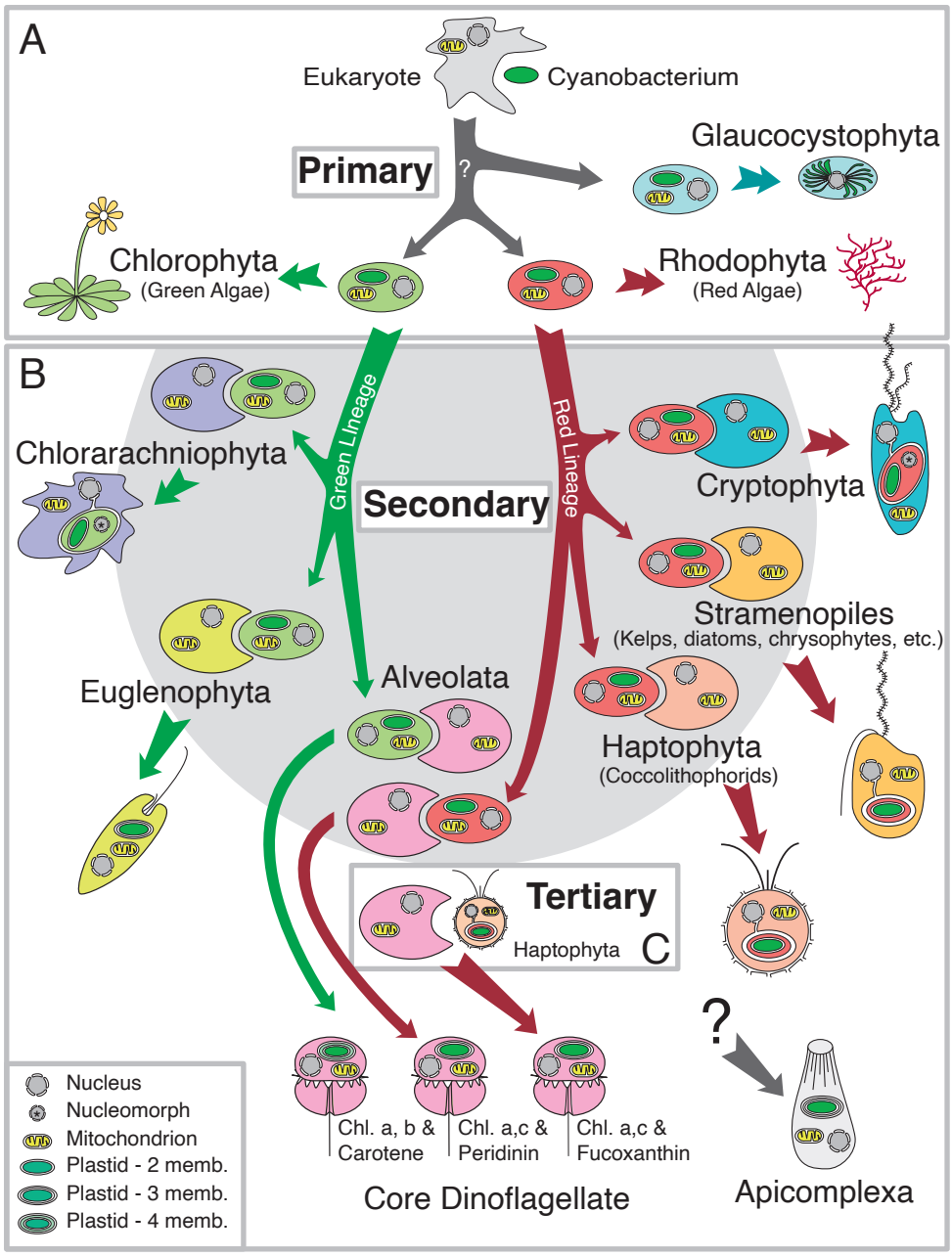



**Figure I-7: Illustration of multiple symbiotic engulfments in multiple algal lineages** Shown is an illustration of how multiple symbiotic engulfments of photosynthetic algae can occur, modified from Delwiche, 1999. The red and green arrows depict the red and green lineages of algal chloroplast based on pigment and gene content. The resultant multiple membrane structure from the serial engulfment of eukaryotic algae are shown in the lower left image.

Still, the abundance of plastid lipids and strong links with photosynthesis indicates that lipid synthesis was acquired with the plastid during endosymbiosis. Additionally, core dinoflagellates possess several 4-methyl and 4-desmethyl steroids in addition to cholesterol, some of which are also found in haptophytes (Mansour et

al., 1999), furthering the hypothesis that a significant amount of biosynthetic machinery in core dinoflagellates is plastid derived. The 18 carbon fatty acids in core dinoflagellates are highly unsaturated, including EPA 20:5 (n-3) and DHA 22:6(n-3), without an abundance of the mono-unsaturated fatty acids commonly found in plants such as oleic acid (18:1), linoleic acid (18:2), and α-linolenic acid (18:3). These are primarily formed by elongases and desaturases in two separate pathways, a cytosolic and a chloroplast pathway (Ohlrogge & Jaworski, 1997), and there is some evidence for the presence of these enzymes in dinoflagellates (Guo, Wang, Liu, & Li, 2021) although with an unknown origin.

Dinoflagellates appear to have acquired both de-novo lipid as well as non-cholesterol sterol synthesis during the transition from syndinian to core dinoflagellates. While the exact origin of de-novo lipid synthesis remains unclear, the similarity of dinoflagellate sterols to haptophyte sterols points to a plastid origin for the biosynthetic machinery. Although these have likely been transferred to the nucleus (Zhang, Green, & Cavalier-Smith, 1999), it is necessary to consider their origin when thinking about functionality and the possibility of novel functions for these genes.

### Toxins and polyunsaturated fatty acids in dinoflagellates

Dinoflagellates are members of the SAR group of protists (Cavalier-Smith, 1998; Patterson, 1999), SAR being an acronym for **S**tramenopiles, **A**lveolates, and **R**hizaria. Alveolates include ciliates, dinoflagellates as well as apicomplexans and chromerids that are frequently parasitic except most core dinoflagellates and a few chromerid species that are photosynthetic, although apicomplexans have a relic plastid (Sato, 2011). Similarly, stramenopiles include several photosynthetic members, such as diatoms, brown algae as well as several fungal-like non-photosynthetic members like the oomycetes and Labyrinthulomycetes. All photosynthetic stramenopiles have evidence of a secondary endosymbiosis with multiple membranes around their chloroplasts, similar to core dinoflagellates(Sato, 2020), indicating that their chloroplast was a free-living photosynthetic eukaryote, likely a red algal derivative (Janouskovec et al., 2010; Delwiche, 1999). Many stramenopiles, photosynthetic and non-photosynthetic alike, produce polyunsaturated fatty acids such as eicosapentaenoic acid 20:5(n-3), docosahexaenoic acid 22:6(n-3) and arachidonic acid 20:4(n-3). Well studied groups that make polyunsaturated fatty acids include diatoms (Dunstan, Volkman, Barrett, Leroi, & Jeffrey, 1993) and haptophytes (Guihéneuf & Stengel, 2013) that are both photosynthetic as well as thraustochytrids that are non-photosynthetic (Raghukumar, 2002), although diatoms frequently produce aldehyde forms that theoretically reduce grazing (Yi, Xu, Di, Brynjolfsson, & Fu, 2017). The relative amounts and types of long-chain polyunsaturated fats in stramenopiles are very similar to those in core dinoflagellates (Peltomaa, Hällfors, & Taipale, 2019) and, given the haptophyte origin of the dinoflagellate plastid, may share a common ancestry across these lineages.

Carbon-13 labeling in haptophytes and thraustochytrids has shown that these long- chain polyunsaturated fatty acids are predominantly made by polyketide

synthesis (PKS) directly instead of by synthesizing lipids followed by elongation and desaturation (Remize et al., 2021; Wang et al., 2020). Polyketide synthesis is nearly identical to lipid synthesis and utilizes the same biochemistry (Bentley & Bennett, 1999). The sequential incorporation of acetate in lipid synthesis is by definition a polyketide but in lipid synthesis the ketones are all fully reduced to a saturated acyl chain in an iterative process. In the broader scope of polyketide synthesis, the saturation of each incorporated moiety is variable, as is the substrate itself, and the process is modular rather than iterative as in lipid synthesis, allowing for an incredible range of resultant products. The thraustochytrid *Aurantiochytrium limacinum* shows three PKS gene clusters that impact docosahexaenoic acid production (Ren et al., 2018; Liu et al., 2018). These encode ketosynthases that incorporate both acetate and malonate as well as reductases and dehydratases that produce the sequential saturated and unsaturated regions. Core dinoflagellates possess these same genes (Bachvaroff, Williams, Jagus, & Place, 2015; John et al., 2008; Kohli et al., 2016; Van Dolah, Kohli, Morey, & Murray, 2017) and have, at least theoretically, the synthetic machinery to make polyunsaturated fatty acids in the same manner as stramenopiles. Additionally, the syndinian dinoflagellate *Hematodinium* sp. appears to have a transcript with similar functionality (Gornik et al., 2015), although its evolutionary origin is unknown.

Things would be simple if dinoflagellates only made polyunsaturated fatty acids using PKS machinery. In fact, many natural products have been described in dinoflagellates, a natural product herein defined as any free non-polymeric molecule that is biosynthesized. Just as with polyunsaturated fatty acids, the natural products likely made by polyketide synthesis were described using carbon isotope labeling (Fukatsu et al., 2007; Ishida et al., 1995; Meng, Van Wagoner, Misner, Tomas, & Wright, 2010; Rasmussen et al., 2017; Sasaki et al., 1996; Seki, Satake, Mackenzie, Kaspar, & Yasumoto, 1995; Van Wagoner et al., 2008; Van Wagoner et al., 2010; Kobayashi, 2008) (Figure I-8). Unlike polyunsaturated fatty acids, these molecules are

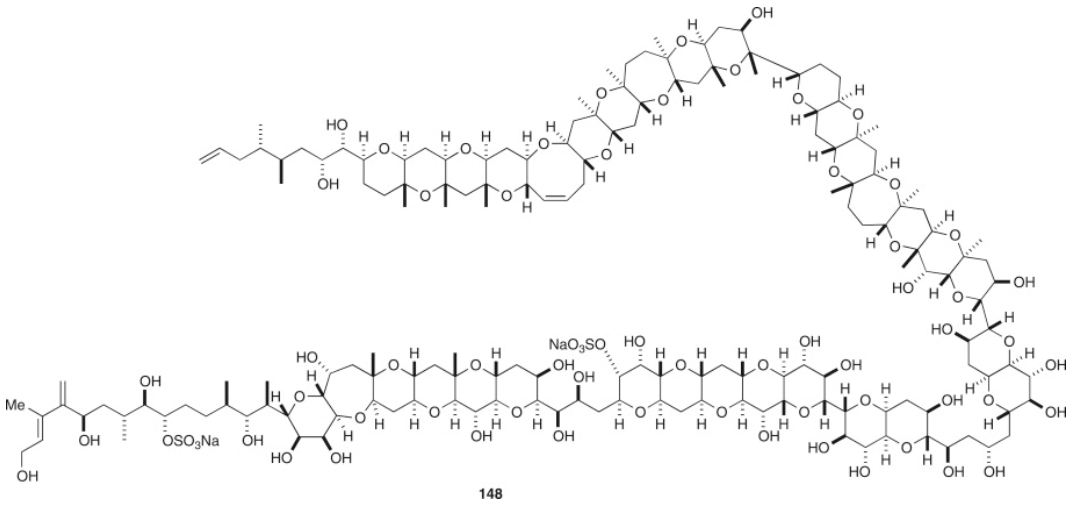



**Figure I-8: Two-dimensional structure of maitotoxin**
An example dinoflagellate toxin is shown as a two dimensional stick representation with bonds shown an black lines, carbons shown as line intersections, and other atoms represented by their IUPAC abbreviations. This particular molecule is maitotoxin from the core dinoflagellate *Gambierdiscus toxicus* showing the ether bonds resulting from hydroxyl groups bonding to carbon molecules.

composed entirely of acetate and not malonate, including a few small carboxylic acids such as glycolate and glycine being documented exceptions. Many of these natural products pose a threat to human and environmental health (Anderson, 1994; Walsh et al., 2015; Wang, 2008; Withers, 1982) and, as such have been labeled toxins. Their ecological roles are less well understood. Most of these compounds are retained in the cell except karlotoxin from *Karlodinium veneficum*, which acts in prey capture and predator avoidance (Adolf, Krupatkina, Bachvaroff, & Place, 2007; Sheng, Malkiel, Katz, Adolf, & Place, 2010) (Figure I-9).

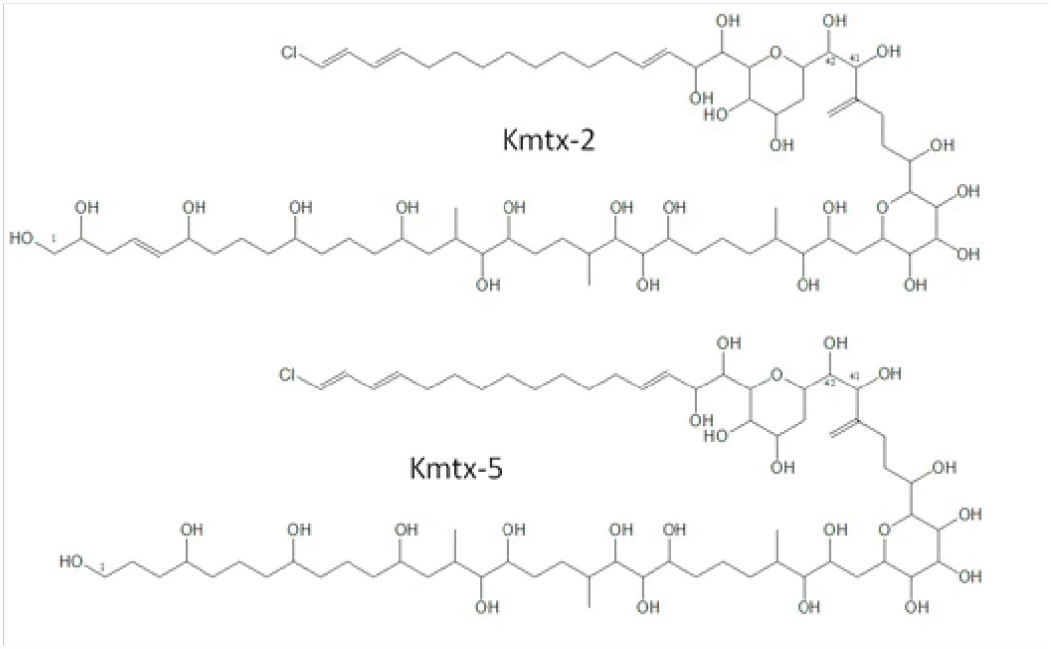



**Figure I-9: Two-dimensional structure of karlotoxins 2 and 5**
Two-dimensional representations of karlotoxin 2 and 5 from *Karlodinium veneficum* are shown with bonds represented by black lines, carbon represented by line vertices, and other atoms given as their IUPAC abbreviation from Van Wagoner *et al*. 2014. This representation shows the hairpin-like configuration theorized to bind cholesterol within a membrane and create a pore.

Brevetoxin from *Karenia brevis* (Baden, 1989) (Figure I-10) has been

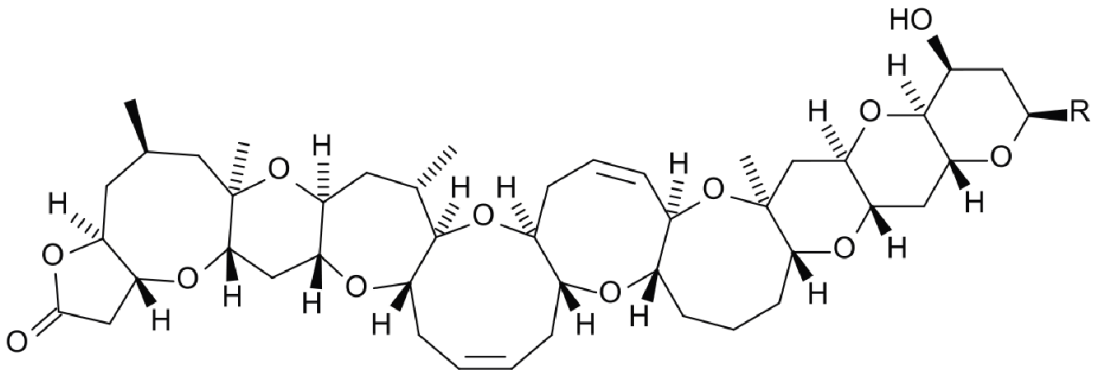


**Figure I-10: Two-dimensional structure of brevetoxin**
A two-dimensional representation of brevetoxin from *Karenia brevis* is shown with bonds represented by black lines, carbon represented by line vertices, and other atoms given as their IUPAC abbreviation described in Baden, 1989. In this drawing the "R" on the far right represents variable groups that distinguish brevetoxin sub-types.

associated with intracellular redox state and carotenoid remodeling (Colon et al., 2021; Chen et al., 2018), but otherwise, the biological roles of these compounds

remain a mystery. Despite this, there is the potential to exploit these compounds for human use since they can bind cholesterol to form pores and block a variety of ion channels (Javed, Qadir, Janbaz, & Ali, 2011; Wang, 2008; Place, Bai, Kim, Sengco, & Wayne Coats, 2009). Unfortunately, their redundant but complex structure makes them difficult or impossible to synthesize in the lab.

Dinoflagellates produce polyunsaturated fatty acids and other often harmful polyketides that have therapeutic uses in humans. The acquisition of the biosynthetic pathways likely occurred or was expanded during the transition from syndinian to core dinoflagellate lineages. Much of the biosynthetic machinery for making all of these natural products, including polyunsaturated fats, are similar, identical, or even overlapping. This makes functional assignments difficult and a necessary area for further study.

### Natural product synthesis in bacterial and fungal models

Many bacteria and fungi make various natural products and have been studied extensively. It could be argued that penicillin was the first natural product (Fleming, 2001), and antibiotics remain a common driver behind advancements in natural product research (Felnagle et al., 2008; Gomes, Schuch, & de Macedo Lemos, 2013). These advancements have been facilitated by the modular nature of the biosynthetic pathways, both in their biochemistry and their gene arrangement (Jensen, 2016; Korman, Ames, & (Sheryl) Tsai, 2010). Generally speaking, Cis-acting elements occur tandemly in the genome, frequently in their order of operation, although trans-acting elements are not uncommon and sometimes bridge multiple pathways for a final product (Fewer et al., 2007; Gurney & Thomas, 2011; Wang, Fewer, Holm, Rouhiainen, & Sivonen, 2014; Khosla, Kapur, & Cane, 2009). One major distinction between bacterial and fungal natural product synthesis is that in fungi, a single multidomain protein often acts as a single module, whereas in bacteria, individual proteins for each domain come together to form a biosynthetic module as an enzymatic complex. These are referred to as type I and type II systems, respectively, and sometimes the trans-acting units in eukaryotes are referred to as type II since a single domain is recruited to a larger complex (Khosla, 2009). This has led to the idea that this nomenclature is phylogenetically informative and separates prokaryotes from eukaryotes. The nomenclature has been conflated with taxonomy since, relatively speaking, the domains of bacteria and fungi studied for natural product synthesis have been extremely narrow and likely do not represent the full diversity of synthesis approaches. While this is coincidentally true for model organisms in this field, in dinoflagellates, both single-domain and multi-domain transcripts have been observed that are similar to both eukaryotes and prokaryotes (Bachvaroff, Williams, Jagus, & Place, 2015; Kohli et al., 2015; Kohli et al., 2017; Van Dolah et al., 2017; Van Dolah et al., 2020). This reduces the information content of the standard nomenclature, which is not predictive for marine protists in general, given the evolutionary distance between them and model systems. Instead, the focus will be placed on the actual domains within an mRNA transcript and whether that transcript has a single domain or multiple domains.

Another common nomenclature when referring to natural product synthesis in PKS versus NRPS. These stand for **P**oly**K**etide **S**ynthase and **N**on-**R**ibosomal **P**eptide **S**ynthase, respectively. They are differentiated by the mechanism and substrate addition. In a PKS system, a small carboxylic acid, usually acetate, is added by a ketosynthase. In this system, the ketosynthase provides the specificity for the specific carboxylic acid being added. The process is driven by an ATP dependent enzyme called acetyl CoA carboxylase that carboxylates the acetyl CoA to malonyl CoA resulting in the release of carbon dioxide when the acetate is added to the growing product by a Claisen condensation (Tong, 2005). The acetyl CoA carboxylase enzyme functions identically during fat and natural product synthesis and dinoflagellates have two copies, a host isoform likely expressed in the cytosol and a plastid copy likely acting in the chloroplast (Haq, Bachvaroff, & Place, 2017). In an NRPS system, a carboxylic acid is also added, frequently an amino acid, resulting in the peptide bond chemically identical to ribosomal protein synthesis from which the NRPS nomenclature arises. Specificity is provided by the adenylation domain that binds the molecule to be added, but the catalysis is provided by a condensation domain that forms the peptide bond and releases water. This process is also ATP driven, but unlike peptide bonds, in the ribosome, the bond can form with any carboxylic acid in the molecule, such as in the R group of an amino acid, making the adenylation domain responsible for substrate specificity as well as the left-right orientation. One example of this is microcystin, where the R group of the glutamic acid forms a peptide bond with the following amino acid (Allender et al., 2009) (Figure I-11). This nomenclature is also sometimes conflated with phylogeny,

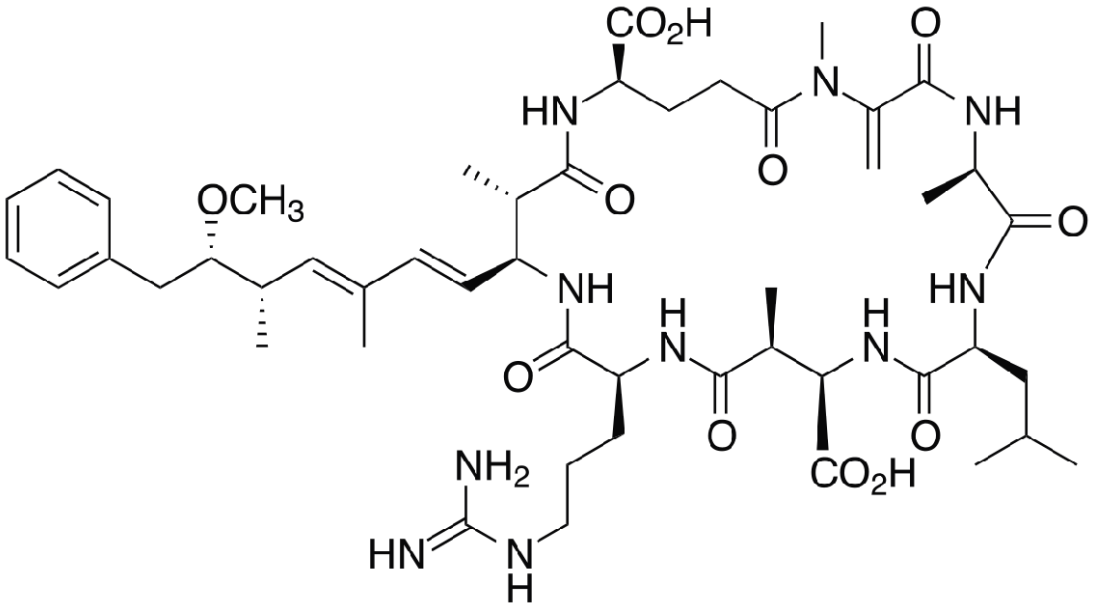



**Figure I-11: Two-dimensional structure of microcystin-LR**
A two-dimensional representation of *Microcystis aeruginosa* is shown with bonds represented by black lines, carbon represented by line vertices, and other atoms given as their IUP AC abbreviation. Each amino acid is specified by an adenylation and then incorporated by a condensation domain. The amino acid on the far left, referred to as ADDA, is first generated by polyketide synthases before incorporation into microcystin.

but microcystin is an example of a hybrid system with both PKS and NRPS elements. In this case, one of the amino acids is generated via a PKS system, followed by incorporation into the microcystin molecule by an NRPS (Tillett et al., 2000). There are many other examples of hybrid systems (Du, Sánchez, & Shen, 2001; Franke, Ishida, & Hertweck, 2012; Kevany, Rasko, & Thomas, 2009), making this nomenclature useful when talking about the potential substrates incorporated into a natural product but not when isolating biosynthetic pathways. Dinoflagellates possess ketosynthases as well as adenylation and condensation domains, sometimes in the same transcripts (Bachvaroff, Williams, Jagus, & Place, 2015), reiterating the need to focus on the domains content of mRNA transcripts and not rely on a global nomenclature with poor predictive power.

Despite the strangeness of their biology, dinoflagellate transcriptomes have revealed that they possess all of the ingredients necessary to synthesize the polyunsaturated fatty acids and toxins that have been described using (all) the pathways described in bacteria and fungi (Kohli et al., 2015; Snyder et al., 2003; Van Dolah et al., 2020). Specifically, what's required is 1) a carrier domain, often called a thiolation domain because a free thiol is created by adding the phosphopantetheinate arm of Coenzyme A to this domain giving a labile group to add and remove each portion of the molecule during synthesis (Lambalot et al., 1996); 2) ketosynthase, adenylation and condensation domains that specify and attach substrate onto the thiolation domain; 3) ketoreductase, dehydratase, and enoyl reductase domains that can reduce the ketone groups to hydroxyl, double, and single bonds, respectively; and 4) acyltransferase and thioesterase domains that remove molecules from the thiolation domains and either move them to another substrate as is the case for acyltransferases or release the molecule as with thioesterases. There are some additional modifications that can occur in dinoflagellates, such as methylation, carbon deletion, or ether ring formation (Van Wagoner, Satake, & Wright, 2014), that likely proceed following the synthesis of the backbone molecule. Altogether this means that we know the types of molecules that dinoflagellates make, their relevance to humans and ecology, as well as the genes that would be used to make them. The questions are which genes are responsible for which compound and how?

### *Amphidinium carterae* as a model for toxin synthesis in Dinoflagellate

The ability for dinoflagellates to synthesize natural products occurred, or at least were enhanced, following the transition from syndinians to core dinoflagellates. Thus, to understand these biosynthetic processes and their evolution, it makes sense to study them in an organism as closely related to the common ancestor of core dinoflagellates as possible. *Oxyrrhis marina* is the most basal core dinoflagellate, but it is not known to make toxins, and its biology is not representative of other core dinoflagellates (Gao & Li, 1986; Montagnes et al., 2011). The next most basal core dinoflagellate is *Amphidinium carterae* (Hulbert) (Figure I-12), which

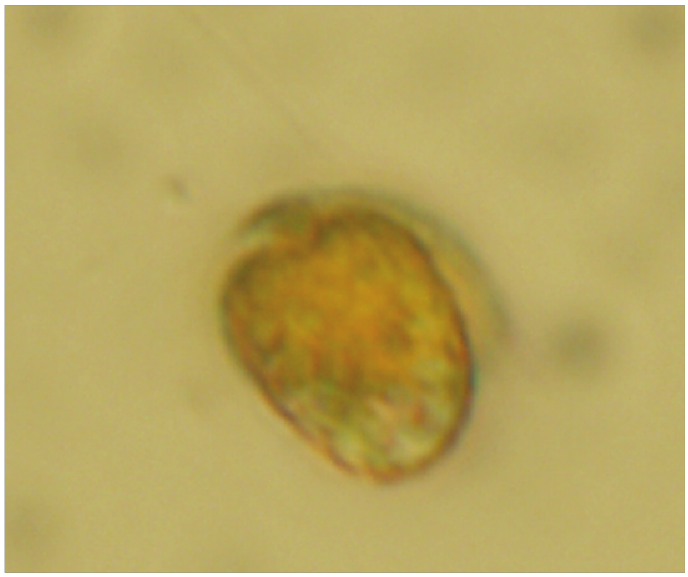


**Figure I-12: Light micrograph of *Amphidinium carterae***
A light micrograph of *Amphidinium carterae* strain NCMA 1314 is shown. The characteristic "golden" color of the peridinin chloroplast is evident as well as one of the flagella extending from the top and curving around the right side.

is similar to many other core dinoflagellates except thecal plates that are lacking in all members of the Gymnodiniales (Bachvaroff et al., 2014). *Amphidinium carterae* also has a relatively small haploid genome size (3.4 picograms) (Holm-Hansen, 1969) compared to other dinoflagellates indicating that gene duplication may be less severe in this species. For example, the essential translation initiation factor 4E is present as eight copies in *Amphidinium carterae* compared to an average copy number of eleven in core dinoflagellates (Jones et al., 2015). *A. carterae* is photosynthetic, has a peridinin containing plastid and makes the hemolytic toxin amphidinol (Houdai et al., 2001; Meng et al., 2010) as well as several other likely partial forms termed amphidinolides (Kobayashi, 2008). Amphidinol is similar in structure to karlotoxin and karmitoxin (Rasmussen et al., 2017; Van Wagoner et al., 2010), containing two ether rings and maybe an evolutionary precursor to the polyether ring toxins (Ishida et al., 1995; Macpherson, Burton, LeBlanc, Walter, & Wright, 2003; Satake et al., 1997; Paz et al., 2008). Most importantly, *A. carterae* can be grown at relatively high densities in axenic culture (Liu, Place, & Jagus, 2017), is readily available from culture collections, has a relatively small diploid genome of around 3.4 picograms (LaJeunesse, Lambert, Andersen, Coffroth, & Galbraith, 2005a) and a typical peridinin plastid. Combined with a robust transcriptome (Lauritano et al., 2017) and the annotation of several PKS/NRPS genes (Bachvaroff, Williams, Jagus, & Place, 2015), *A. carterae* is a sensible candidate for the characterization of genes that may be involved in natural product synthesis.

## Overview of experimental goals and rationale

In order to understand a complicated yet biochemically redundant process like natural product synthesis, we need to be able to isolate the participants for our product of interest, separating the wheat from the chaff. Thus, the goal is to develop methods that will allow for identifying specific interacting partners in natural product synthesis and assign a putative function. The following chapters will address three sequential avenues to achieve this goal. The first chapter focuses on cataloging the genes and domains that may participate in synthesizing these natural products and selecting candidate genes for further study (Williams, Bachvaroff, & Place, Evolutionary Biology, 2021). This is a necessary first step to address the copy number issues. While certain genes may be fundamentally important to biosynthesis, a high copy number may make their study intractable. To the largest extent, possible candidates should be chosen by their ability to be biochemically validated and the strength of current functional predictions. The second chapter attempts to characterize the biology of the candidate genes (Williams, Bachvaroff & Place 2020, Microorganisms). This will inform further analyses that may not occur in-situ by giving downstream results a physiological context. Methods employed should be a mix of direct observations as well sequence-based predictions when reasonable to inform downstream experiments. The third chapter will validate these predictions by testing the interactions of candidates in a heterologous system (Williams, Bachvaroff & Place 2022, Microorganisms). The underlying rationale is that proteins that readily interact are more likely to participate in the same biochemical pathway as biosynthetic partners. One important goal is to distinguish the synthesis of toxins and polyunsaturated fats from the synthesis of saturated lipids, which use many of the same genes. Another important goal is to see how these proteins behave compared to model systems in bacteria and fungi to establish a baseline expectation of function for future studies. Finally, it is worth establishing protocols for characterizing dinoflagellate genes using an expression in a heterologous system and identifying potential pitfalls for future work.

# Chapter 1: A Global Approach to Estimating the Abundance and Duplication of Polyketide Synthase Domains in Dinoflagellates

*Abstract*

Many dinoflagellate species make toxins in a myriad of different molecular configurations but the underlying chemistry in all cases is presumably via modular synthases, primarily polyketide synthases. In many organisms modular synthases occur as discrete synthetic genes or domains within a gene that act in coordination thus forming a module that produces a particular fragment of a natural product. The modules usually occur in tandem as gene clusters with a syntenic or operonic arrangement that is often predictive of the resultant structure. Dinoflagellate genomes, however, are notoriously complex with individual genes present in many tandem repeats and very few synthetic modules occurring as gene clusters, unlike what has been seen in bacteria and fungi. However, modular synthesis requires a free thiol group called a thiolation domain that acts as a carrier for sequential synthesis. We scanned 47 dinoflagellate transcriptomes for 23 modular synthase domain models and compared their abundance among ten orders of dinoflagellates as well as their co-occurrence with thiolation domains. The total count of domain types was quite large with over thirty-thousand identified, twenty-nine thousand of which were in the core dinoflagellates. Although there were no specific trends in domain abundance associated with types of toxins, there were clear lineage specific differences. The Gymnodiniales, makers of long polyketide toxins such as brevetoxin and karlotoxin, had a high relative abundance of thiolation domains as well as multiple thiolation domains within a single transcript. Orders such as the Gonyaulacales, makers of small polyketides such as spirolides, had fewer thiolation domains but a relative increase in the number of acyl transferases. Unique to the core dinoflagellates, however, were thiolation domains occurring alongside tetratricopeptide repeats that facilitate protein-protein interactions, especially hexa and hepta-repeats, that may explain the scaffolding required for synthetic complexes capable of making large toxins. Clustering analysis for each type of domain was also used to discern possible origins of duplication for the multitude of single domain transcripts. Single domain transcripts frequently clustered with synonymous domains from multi-domain transcripts such as the BurA and ZmaK like genes as well as the multi-ketosynthase genes, sometimes with a large degree of apparent gene duplication, while fatty acid synthesis genes formed distinct clusters. Surprisingly the acyltransferases and ketoreductases involved in fatty acid synthesis (FabD and FabG, respectively) were found in very large clusters indicating an unprecedented degree of gene duplication

for these genes. These results demonstrate a complex evolutionary history of core dinoflagellate modular synthases with domain specific duplications throughout the lineage as well as clues to how large protein complexes can be assembled to synthesize the largest natural products known.

## *Introduction*

Dinoflagellates are unicellular aquatic eukaryotes with an interesting and complicated evolutionary history (Bachvaroff et al., 2014; Janouškovec et al., 2017). Generally speaking, they can be divided into two main groupings with the heterotrophic, often parasitic syndiniales at the base of the dinoflagellate lineage and the often mixotrophic "core dinoflagellates" extending out into the distal branches (Janouškovec et al., 2017). The core dinoflagellates have a chloroplast or evidence of a lost chloroplast with multiple symbiotic events occurring throughout the lineages (Keeling, 2010; Schnepf & ElbräChter, 1999). Although many core dinoflagellates are mixotrophic (Jacobson & Anderson, 1996; Jeong et al., 2010), the majority of dinoflagellate "algae" that form harmful algal blooms are photosynthetic, *Noctiluca scintilans* (Macartney) being the exception. Toxic dinoflagellates are exclusively photosynthetic and there is evidence that toxin synthesis may initiate in the chloroplast (Monroe, Johnson, Wang, Pierce, & Van Dolah, 2010; Van Dolah et al., 2013), indicating a potential relationship between photosynthesis and natural product synthesis in dinoflagellates. *Amphidinium carterae* (Hulbert) is a basal, photosynthetic dinoflagellate that makes the toxin amphidinol as well as many derivatives termed amphidinolides (Kobayashi, 2008; Meng et al., 2010), indicating that the acquisition of a plastid and toxicity are early events in the evolution of the core dinoflagellates. Many dinoflagellate toxins pose human health concerns by a variety of mechanisms (Wang, 2008) as well as ecological and trophic impacts (Sheng et al., 2010; Van Wagoner et al., 2010).

The toxins themselves are almost universally polyketides, *i.e.* they are formed from sequentially added acetate subunits that are modified prior to addition of the next acetate subunit (Bentley & Bennett, 1999). The workhouse enzymatic domain in the synthesis of polyketides is the ketosynthase (KS) domain, a condensation domain that incorporates malonyl-CoA into an existing acyl chain as acetate with the release of CO2 driving the reaction (Jenke-Kodama & Dittmann, 2009; Khosla, 2009). Analogously non-ribosomal peptide synthases also perform a condensation reaction of carboxylic acids, often an amino acids using a condensation domain but with substrate specificity provided by an adenylation domain (Izoré & Cryle, 2018; Rausch, Hoof, Weber, Wohlleben, & Huson, 2007). The enzymes that incorporate each building block work with downstream modifying domains to form synthetic **modules** that create complex biomolecules and are responsible for many known naturally occurring compounds including antibiotics (Gurney & Thomas, 2011; Lim et al., 2009; McDaniel et al., 1999).

Labeling studies have shown that dinoflagellate toxins exclusively incorporate acetate from malonyl-CoA (Houdai et al., 2001; Lee, Qin, Nakanishi, & Zagorski, 1989; Macpherson et al., 2003; Meng et al., 2010; Wright, Hu, McLachlan, Needham, &

Walter, 1996), unlike bacteria that often incorporate propionate or butyrate (Moore & Hertweck, 2002). Toxins often initiate from or are occasionally extended by small amino acids like glycine or other carboxylic acids like glycolate (Macpherson et al., 2003; Rasmussen et al., 2017). There is also evidence for alkylation by methionine or acetate as well as side-by-side 'alpha' carbons from acetate explained by the deletion of carbon by a theorized Favorskii rearrangement removing the beta carbon from one acetate (Van Wagoner et al., 2014). Toxins range in complexity and size from the 31 carbon Gymnodimine (Seki et al., 1995) to the 164 carbon maitotoxin that has 98 stereocenters (Sasaki et al., 1996).

In spite of their complexity, synthesis of the backbone of dinoflagellate toxins utilizes the same core machinery as lipid synthesis. Lipids as a secondary metabolite are differentiated from natural products in that they are fully saturated, highly regulated, and usually transported to the chloroplast, mitochondrion, and cytosol during both synthesis and degradation (Buhman, Chen, & Farese, 2001; Marechal, Block, Dorne, Douce, & Joyard, 1997; Tatsuta, Scharwey, & Langer, 2014). Thus, there is frequently a segregation of genes, phylogenetically and physically, involved in lipid synthesis from those involved in secondary metabolite synthesis, including in dinoflagellates (Kohli et al., 2016). In terms of acetate incorporation, all dinoflagellate toxins and lipids rely on the aforementioned ketosynthase domains along with several biologically universal modification domains: ketoreductases (KRs), dehydratases (DHs) and enoyl reductases (ERs) to form a sequentially reduced backbone structure and acyl transferases (ATs) and thioesterases (TEs) to move and terminate growing acyl chains. These enzymatic domains interact with the substrate and each other via a reaction center created by transferring the phosphopantetheinyl arm of coenzyme A onto a carrier protein (Beld, Sonnenschein, Vickery, Noel, & Burkart, 2014; Wang et al., 2014).

One key difference between lipid and other secondary metabolite synthesis is that lipid synthesis is iterative, utilizing a single carrier protein called the acyl carrier protein while natural products are made with multiple modules with a homologous thiolation carrier domain . Whether the particular chemistry of each module is a PKS, an NRPS, or a hybrid system; a thiolation domain is the reaction center for all of these modular synthases. Likewise, a thiolation domain is the reaction center for each **module** involved in toxin synthesis in dinoflagellates. This is useful when dealing with dinoflagellates since the type I multi-domain polyketide synthases and non-ribosomal peptide synthases found in fungi (Schümann & Hertweck, 2006) and usually associated with eukaryotes are relatively uncommon in dinoflagellate transcriptomes with most described transcripts containing one or rarely a few domains that would have to be combined into a multi-enzyme **synthetic complex** (Van Dolah et al., 2017), similar to the type II polyketide synthases and NRPS usually found in prokaryotes (Hertweck, Luzhetskyy, Rebets, & Bechthold, 2007; Izoré & Cryle, 2018). This is not surprising since dinoflagellates often encode genes as tandem repeats of gene copies rather than gene clusters of common metabolic function (Bachvaroff & Place, 2008), but this also makes phylogenetic reconstruction difficult even for single domains due to a high copy number of very similar sequences.

The exceptions to the multitude of single domain transcripts in dinoflagellates are several multi-domain genes that have conserved domain arrangement and sequence. Two of these are the BurA and ZmaK-like genes (Bachvaroff, Williams, Jagus, & Place, 2015) each contain both adenylation and ketosynthase domains in what appears to be a single hybrid PKS NRPS module. There is also a multi-module gene usually containing at least three consecutive ketosynthase-containing modules, here referred to as the triple KS (Van Dolah et al., 2017; Van Dolah et al., 2020). Phylogenies of dinoflagellate modular synthase domains usually form a robust set of dinoflagellate clades but with poor support placing these clades among eukaryotic outgroups, as well as no obvious reflection of relationships within core dinoflagellates (Beedessee et al., 2019; Kohli et al., 2016; Meyer et al., 2015). This not only reveals a gap in annotated sequences that can function as outgroups to dinoflagellates but also indicates that at least some of these modular synthases are likely of bacterial origin, specifically BurA and ZmaK, which have only been described in prokaryotes and seem to have been transferred in their entirety (Franke et al., 2012; Kevany et al., 2009). Thus, with the exceptions of the conserved fatty acid biosynthetic genes and the corresponding acyl carrier protein, phylogenetic comparisons to model eukaryotes or prokaryotes are generally uninformative when trying to deduce the roles of dinoflagellate modular synthases in toxin production. Likewise, the traditional nomenclature of polyketide synthases that relies on single versus multi-domain and prokaryote versus eukaryote fails to describe the domains in dinoflagellates in a useful manner.

The primary aim of this study was to survey genes that may be involved in dinoflagellate natural product synthesis, specifically toxins but including lipid synthesis, without prejudice from prokaryotes or distantly related model eukaryote terminology. *Amphidinium carterae* was used as a model because it is a basal toxic dinoflagellate (Janouškovec et al., 2017) and has the BurA- and ZmaK-like genes as well as the triple KS gene in their apparent entirety and single copy (Bachvaroff, Williams, Jagus, & Place, 2015; Van Dolah et al., 2017). The domains selected were taken from these previously annotated multi-domain dinoflagellate transcripts using the following strategy 1) Use the *A. carterae* to collect similar domains from the *A. carterae* transcriptome, 2) Make a hidden Markov Model (HMM), 3) Use the HMM to retrieve domains from all available dinoflagellate transcriptomes, and 4) Enumerate and cluster resultant domains to functionally bin them.

This resulted in several unexpected discoveries such as a large number of adenylation domains seemingly without the traditional condensation domains as well as scaffolding domains associated with specific synthetic domains. This global approach was also able to describe the relative copy number of each synthetic domain revealing several atypical relationships. One example is a large number of enoyl reductases compared to dehydratases, which is very strange since enoyl reductases theoretically act downstream and should be less abundant than dehydratases. The second portion of this survey was to place the retrieved domains into theoretical functional bins based on sequence similarity using a method that is not hampered by gene duplication and horizontal gene transfer. Although many of these synthetic domains can and have been given hypothetical function using phylogenetic inference with model systems as outgroups, the results presented here demonstrate many novel sequence clusters that

are difficult to resolve phylogenetically as well as some very atypical gene expansions including acyl transferases and ketoreductases involved in lipid synthesis that were largely overlooked in previous studies. The results of this study demonstrate another way in which dinoflagellates defy the paradigms established by model systems, in this case in terms of the mechanisms of natural product (toxin) synthesis and are presented here as a framework to be used in future biochemical experiments to validate the hypothetical functions of PKS and NRPS genes in dinoflagellates.

## *Materials and Methods*

### Transcriptome preparation and analysis

A total of 61 initial transcriptomes were selected for domain searches with the majority of dinoflagellate transcriptomes taken from the CAMERA database originally published in (http://camera.calit2.net/mmetsp/list.php, (Sun et al., 2011) and now hosted at https://www.imicrobe.us/#/projects/104. The NCBI project #PRJNA231566 was assembled using Trinity v.2.0.2. In addition, data for cultures of *Karenia brevis* (C.C. Davis), *Karlodinium veneficum* (D. Ballantine), and *Akashiwo sanguinea* (K. Hirasaka) that were infected with the syndinian parasite of the genus *Amoebophyra* were collected from previous phylogenetic studies (Bachvaroff, Handy, Place, & Delwiche, 2011; Bachvaroff et al., 2014; Williams et al., 2017). For *A sanguinea*, the transcriptome was done with and without infection and for the *K. veneficum* parasite there is a genome available for comparison (Bachvaroff, 2019). In addition to these transcriptomes, the deep sequencing transcriptomes (using Hi-Seq) for *K. brevis* (Van Dolah et al., 2017), and two *Gambierdiscus* species (Kohli et al., 2017), *G. excentricus* (S. Fraga) and *G. polynesiensis* (Chinain and M. Faust), assembled using CLC (595M, 118M, 884M reads, respectively) were included. Unfortunately, the transcriptomes from the two *Gambierdiscus* species in the transcriptome sequence archive were incomplete with about seventy PKS genes identified in the initial study deposited separately in Genbank. These were added back into the total domain count following domain searches. Each transcriptome was translated in all six frames using a Perl script and Genbank translation table 1 (standard eukaryotic) prior to analysis.

Benchmarking Universal Single-Copy Orthologs (BUSCO) (Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015) was used to determine transcriptome quality using the Eukaryota odb9 dataset (Table 1-1).

**Transcritome BUSCO Scores and Domain Content**

| Species | Combined § | Single§ | Duplicate § | Domains | Domain Types |
|---|---|---|---|---|---|
| Alexandrium andersonii | 33.30% | 25.70% | 7.60% | 1155 | 541 |
| Alexandrium catenella | 68.30% | 54.10% | 14.20% | 1244 | 584 |
| Alexandrium margalefi | 73.00% | 59.10% | 13.90% | 1555 | 801 |

| | | | | | |
|---|---|---|---|---|---|
| Alexandrium minutum | 27.40% | 22.10% | 5.30% | 169 | 91 |
| Alexandrium monilatum | 80.20% | 57.40% | 22.80% | 1292 | 678 |
| Alexandrium tamarense | 86.10% | 49.80% | 36.30% | 1739 | 956 |
| Karlodinium veneficum | 78.90% | 55.80% | 23.10% | 1227 | 638 |
| Amphidinium carterae | 78.60% | 66.70% | 11.90% | 727 | 388 |
| Amphidinium klebsii | 80.20% | 66.70% | 13.50% | 720 | 393 |
| Amphidinium massartii | 77.30% | 67.70% | 9.60% | 614 | 314 |
| Akashiwo sanguinea | 83.10% | 46.50% | 36.60% | 1497 | 848 |
| Azadinium spinosum | 80.20% | 57.80% | 22.40% | 2122 | 1108 |
| Brandtodinium nutriculum | 64.60% | 52.10% | 12.50% | 859 | 420 |
| Ceratium fusus | 81.50% | 60.40% | 21.10% | 1066 | 653 |
| Chromera velia | 54.80% | 47.50% | 7.30% | 104 | 66 |
| Crypthecodinium cohnii | 79.20% | 63.40% | 15.80% | 1231 | 716 |
| Dinophysis acuminata | 71.00% | 53.80% | 17.20% | 1590 | 787 |
| Durinskia baltica | 81.10% | 45.50% | 35.60% | 801 | 387 |
| Gambierdiscus excentricus | 74.30% | 63.70% | 10.60% | 847 | 448 |
| Gyrodinium instriatum | 80.90% | 53.50% | 27.40% | 2110 | 1147 |
| Glenodinium foliaceum | 81.80% | 46.20% | 35.60% | 1078 | 532 |
| Gonyaulax spinifera | 50.90% | 38.00% | 12.90% | 883 | 412 |
| Gambierdiscus polynesiensis | 68.30% | 59.70% | 8.60% | 1378 | 766 |
| Gymnodinium catenatum | 83.20% | 63.70% | 19.50% | 579 | 325 |
| Hematodinium sp. | 77.20% | 33.30% | 43.90% | 724 | 471 |
| Heterocapsa arctica | 64.00% | 51.80% | 12.20% | 797 | 398 |
| Heterocapsa rotundata | 65.70% | 57.10% | 8.60% | 712 | 343 |
| Karenia brevis CLC | 80.90% | 64.70% | 16.20% | 2006 | 1197 |
| Karenia brevis Trinity | 83.50% | 61.40% | 22.10% | 1526 | 939 |
| Kryptoperidinium foliaceum | 84.80% | 36.00% | 48.80% | 1699 | 823 |
| Lingulodinium polyedra | 80.80% | 58.40% | 22.40% | 2407 | 1304 |
| Noctiluca scintilans | 77.50% | 65.00% | 12.50% | 625 | 324 |
| Oxyrrhis marina (LB1974) | 75.20% | 60.70% | 14.50% | 399 | 240 |
| Oxyrrhis marina (unknown) | 79.50% | 62.00% | 17.50% | 461 | 290 |
| Pelagodinium beii | 68.00% | 53.10% | 14.90% | 945 | 462 |
| Peridinium aciculiferum | 78.80% | 58.70% | 20.10% | 846 | 443 |
| Perkinsus chesapeaki | 1.70% | 1.70% | 0.00% | 16 | 9 |
| Perkinsus marinus | 30.00% | 23.10% | 6.90% | 75 | 40 |
| Prorocentrum hoffmanianum | 79.60% | 58.10% | 21.50% | 1178 | 581 |
| Prorocentrum micans | 78.30% | 47.90% | 30.40% | 1210 | 696 |

| | | | | | |
|---|---|---|---|---|---|
| Prorocentrum minimum (1329) | 76.90% | 48.50% | 28.40% | 678 | 370 |
| Prorocentrum minimum (2233) | 39.20% | 30.00% | 9.20% | 276 | 155 |
| Polarella glacialis (2088) | 55.80% | 44.20% | 11.60% | 594 | 298 |
| Protoceratium reticulatum | 68.30% | 54.10% | 14.20% | 1247 | 592 |
| Pyrodinium bahamense | 73.30% | 61.40% | 11.90% | 1138 | 598 |
| Scrippsiella hangoei | 81.90% | 57.10% | 24.80% | 1234 | 724 |
| Scrippsiella hangoei like | 82.50% | 62.00% | 20.50% | 1002 | 532 |
| Scrippsiella trochoidea | 77.80% | 55.40% | 22.40% | 1554 | 906 |
| Symbiodinium (B1) | 80.90% | 70.00% | 10.90% | 676 | 385 |
| Symbiodinium (C1) | 84.80% | 62.00% | 22.80% | 199 | 109 |
| Symbiodinium (C15) | 35.30% | 32.00% | 3.30% | 828 | 418 |
| Symbiodinium (2430) | 57.10% | 49.50% | 7.60% | 568 | 278 |
| Symbiodinium (421) | 66.30% | 30.00% | 36.30% | 1352 | 659 |
| Symbiodinium (D1a) | 46.90% | 28.40% | 18.50% | 558 | 307 |
| Symbiodinium (Mp) | 81.60% | 70.00% | 11.60% | 766 | 368 |
| Symbiodinium (A) | 61.40% | 52.50% | 8.90% | 752 | 381 |
| Triceratium dubium | 41.50% | 32.30% | 9.20% | 183 | 98 |

§ Percentages shown are the fraction of BUSCO genes retrieved by one (Single), multiple (Duplicate), or any number (Combined) of transcripts in each transcriptome

The eukaryote database was chosen over the protist database after initial tests with the protist database gave very low scores (approximately 30% maximum, data not shown). This study is intentionally specific to the "core" dinoflagellates so only closely related outgroups were used including *Perkinsus marinus* (Levine), *Chromera velia* (R.B. Moore *et al.*), and *Triceratium dubium* (Brightwell), as well as the syndinian parasite of crustaceans *Hematodinium* sp. and the two aforementioned syndiniales with dinoflagellate hosts. No pertinent domains were found in the transcriptomes of syndinian parasites with dinoflagellate hosts except for three transcripts from the *K. veneficum* parasite and were thus excluded from further analyses giving a total of forty-six transcriptomes with a BUSCO score 64% or greater that were included in the final tabulations following domain searches. The two *K. brevis* transcriptomes using different assembly methods had similar BUSCO scores and so the Trinity assembled transcriptome was selected for the final tabulations to make comparisons with other transcriptomes, the majority of which were assembled using Trinity, more comparable. *Oxyrrhis* and all outgroup species were given their own taxonomic bin. The forty remaining ingroup transcriptomes were placed into 7 taxonomic bins at approximately the ordinal level including the Gonyaulacales (ten species), the Thoracosphaerales (*Brandtodinium*), the Prorocentrales (three species), the Peridiniales (ten species), the Dinophysiales (two species), the Noctilucales (*Noctiluca*), and the Gymnodiniales (eight species) with the Suessiales (five additional species) as a subgrouping of the Gymnodiniales. The 64%

cutoff was chosen as a natural observed breakpoint for transcriptomes that had a full repertoire of domains relative to other closely related species (Figure S1-1). Some of the outgroup species had lower BUSCO scores (*P. marinus* 30%, *C. velia* 54.8%, *T. dubium* 41.5%) than the 64% cutoff. Although the scores were low, these transcriptomes were included since most of the tabulations are based on ratios and the domain searches successfully recovered transcripts with synthetic modules, *e.g.,* 183 domain hits for *T. dubium* and 104 domain hits for *C. velia*. Also, BUSCO analysis of the *P. marinus* genome (Genbank Bioproject PRJNA12737) yielded a completeness score of 53.3% indicated that the alveolate sequences may not be well represented in the BUSCO database and/or that parasitism has resulted in gene reduction. A lack of sequence representation in the BUSCO database is also supported by maximum BUSCO scores of approximately 80%, even for deeply sequenced transcriptomes showing that the BUSCO scores could be used as a guide but were not quantitative.

### HMM assembly and domain searches

*Amphidinium carterae* (Hulbert) was used to create dinoflagellate specific hidden Markov models (HMMs) of domains from modular synthases. Although many robust models exist for model species, protists in general are poorly sampled and with almost no experimental verification, predictions based on those models are difficult given high pairwise differences. Four transcripts of multi-domain synthases from the *A. carterae* transcriptome were used (Figure 1-1). Each is readily found in other

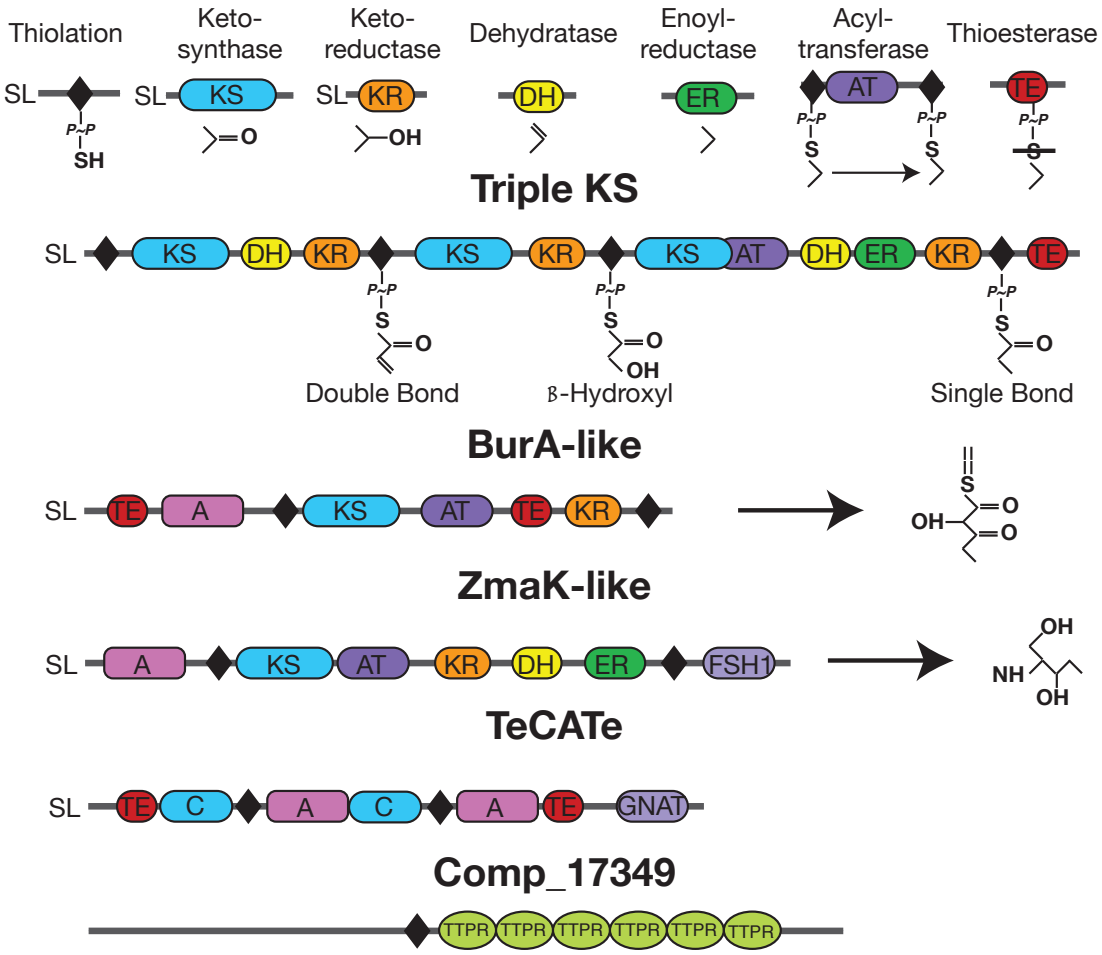



**Figure 1-1: Domain arrangement of *A. carterae* transcripts used in hidden Markov model creation.**
Individual modular synthase domains are shown at the top with example products for their reaction. In addition Adenylation (A), FSH1 serine hydrolases (FSH1), GCN5-associated N-acetyl transferase (GNAT), and tetratrico peptide repeats (TTPR) are shown for the multi-domain transcripts with examples of potential products included. "SL" refers to the dinoflagellate spliced leader sequence and is present if a spliced leader sequence has been verified.

dinoflagellate taxa with the same domain arrangement. The first (comp6001_c0_seq1) is a hybrid PKS / NRPS, the BurA-like gene described in the bacterial genus *Burkholderia* that participates in the synthesis of burkholderic acid (Franke et al., 2012). It has an unusual domain order containing two thioesterase, two thiolation, an adenylation (described by the NCBI conserved domain database as an acyl-CoA ligase), a ketosynthase, a ketoreductase, and an acyltransferase domain. The second (comp26075_c0_seq1) is also a hybrid PKS / NRPS that is most similar on a sequence similarity basis to the ZmaK gene described in *Bacillus cereus* to act in the synthesis of zwittermicin (Kevany et al., 2009). It contains two thiolation, an adenylation, an acyltransferase, a ketosynthase, a ketoreductase, a dehydratase, an enoyl reductase, and a FSH1 serine hydrolase domain. While the BurA-like gene has

the same domain content and arrangement as the genes from *Burkholderia,* the domain arrangement of the ZmaK-like gene is similar but not identical to the ZmaK gene from *B. cereus* making predictions about substrates or function in dinoflagellates unreliable. The third multi-domain transcript is a straightforward multiple ketosynthase-containing set of overlapping transcripts (comp305_c0_seq1, and comp32615_c0_seq1) that have a total of four thiolation domains and three possible modules, each with a ketosynthase domain as well as a ketoreductase; a ketoreductase and a dehydratase; and a ketoreductase, a dehydratase, an enoyl reductase, and a thioesterase. This triple-KS transcript has a ketosynthase in the third module described as an acyltransferase containing ketosynthase by the NCBI conserved domain database. Thus, an acyltransferase may or may not be detected depending on the software used and database queried. A final *A. carterae* transcript (comp14261_c0_seq1) used to make HMMs is herein termed TeCATe due to the flanking thioesterase domains and repeating adenylation and condensation domains as well as a GCN5-associated N-acetyl transferase (GNAT) domain that transfers acetate from acetyl CoA to a substrate (Favrot, Blanchard, & Vergnolle, 2016), but conservation of this sequence in other dinoflagellate species is low. One additional sequence is comp17349_c0_seq1 that contains a thiolation domain and a tetratricopeptide repeat that is used in protein-protein interactions across life in a variety of process and configurations (Zeytuni & Zarivach, 2012). This combination was first described in *K. brevis* (Van Dolah et al., 2017) and was included to determine the prevalence and association of this repeat domain with other modular synthase domains. It is unclear if any of these transcripts participate in toxin synthesis but they are readily identifiable and the domain arrangement of the triple-KS, BurA and ZmaK like genes in *A. carterae* is conserved in other dinoflagellates indicating that the function is also likely conserved.

This resulted in a total of 22 domains for HMM creation with sequence boundaries based on InterPro (Hunter et al., 2009) annotations as implemented in Macvector V16.0.1. These included the adenylation, the ketosynthase, the ketoreductase, and the acyltransferase domain as well as thioesterase domain 1 from BurA; the adenylation, dehydratase, enoyl reductase, and serine hydrolase domains from ZmaK; ketoreductase domain 2, thiolation domain 3, ketosynthase domain 3, dehydratase domain 3, enoyl reductase domain 3, and the thioesterase domain from the triple-KS; adenylation domain 1, thiolation domain 1, both condensation domains, and the GNAT domain from TeCATe; and finally the thiolation and tetratricopeptide repeat domains from comp_17349_c0_seq1 (Figure 1-1). These domains were chosen to provide replicative sampling of each domain across multiple sequences when possible.

The protein translation from the *A. carterae* sequence of each domain was used as the query sequence for a BLAST search across all possible protein translations of the *A. carterae* transcriptome with no cutoff to give as broad a sampling as possible. The aligned region of each BLAST hit was then compiled into a single file for each query domain in fasta format and aligned using Muscle V3.8.31 (Edgar, 2004). These alignments were then each used to generate an *A. carterae* specific hidden Markov model (HMM) for each domain using hmmbuild in the HMMER V3.3 package (Mistry, Finn, Eddy, Bateman, & Punta, 2013). Each HMM

was then compressed with hmmpress and used by hmmsearch with an e-value cutoff of 1e-10 across the protein translations of all 58 transcriptomes with the results given in tab-delimited format for processing. An e-value cutoff was given for the HMM search and not the BLAST search with the assumption that spurious BLAST hits would be represented in the HMM as aligned characters with very low bit scores and that the e-value cutoff in the HMM search would prevent propagation of these errors while maximizing sensitivity. A Perl script was then used to tabulate the data from the HMM search giving a count of each HMM for a given transcript (Table S1-1). The tabulated results were summarized graphically in R V3.3.2 using the GGplot package. For redundant domains the HMM with the highest number of counts for a given transcript was used to maximize sensitivity, *e.g.,* if the three ketosynthase HMMs returned counts of 1, 2, and 1 the transcript would be counted as having two ketosynthase domains. This differentiates the domains that correspond to a specific HMM, from a domain type that equals the functional classification such as "ketosynthase".

### Domain clustering

Protein sequences for each domain were retrieved from the 6-frame translation of each transcriptome using a Perl script and the output from hmmsearch giving the translation frame and position of the alignments for each HMM. Multiple hits within a transcript were indexed out of the maximum number of that domain in the transcript, *e.g.* Ketosynthase 1_3 for ketosynthase domain was the first of a total of three along with the transcript and host indentifiers. The sequence files were dereplicated via a Perl script prior to clustering to remove redundant protein sequences t. The extracted protein sequences were output in fasta format and sequences for adenylation, ketosynthase, thiolation, acyltransferase, and thioesterase domains were each clustered using CLANS (Frickey & Lupas, 2004). This software uses an all by all BLAST search and the subsequent e-values are used as attraction values to group sequences. This is not as robust a method as global alignment with phylogenetic inference for finding relationships but is able to group sequences in three-dimensionsl way and is useful when trying to compare and visualize many very similar sequences. Clusters were visually identified based on a high relative number of internal edges and the sequence names of each node within each cluster were exported to a text file. The sequence list of each cluster was then compared to the master list of domain counts for each transcript to determine the content of each cluster that could be annotated. The vast majority of sequences were single domain transcripts. However, if all of a particular domain from a multi-domain transcript was encompassed by a single cluster then that cluster was labeled based on that multi-domain transcript, *e.g.* if the ketosynthase domain from every BurA transcript was found in a single cluster then that cluster was labeled "BurA".
The thiolation domains were a special case in that almost all of the domains retrieved formed just one cluster. In order to provide resolution, the acyl carrier protein sequences (presumably involved in lipid synthesis) from *A. carterae* (comp649_c0_seq2, comp2819_c0_seq1, and comp3690_c0_seq1) were used as

BLAST queries against the other transcriptomes and a separate HMM search was performed. The thiolation domains from these sequences were then added back into the clustering analysis. This was not necessary for the other domains where either the genes involved in lipid synthesis were retrieved in the initial HMM search or there was sufficient resolution of clusters to make the inclusion of fatty acid biosynthesis genes unnecessary. For the smaller datasets of acyltransferase and thioesterase domains, verification of the clustering results were attempted by maximum likelihood based phylogenetic inference using RAxML (Stamatakis, 2014) using rapid bootstrapping of 100 replicates and seed values of 11111 for both the bootstrapping and parsimony steps.

## *Results*

### BUSCO scores

The scores from the BUSCO analysis ranged from 1.6% for *P chesapeaki* to 86.1% for *A. tamarense* (Table 1-1). There was also frequent duplication with up to 48.8% of the orthologs used for testing present in multiple copies in *Kryptoperidinium foliaceum* (F. Stein). Despite deep sequencing of several of the transcriptomes, the highest BUSCO score would not be considered a complete transcriptome, indicating that many of the "common" eukaryotic orthologs are not present or were not detected. Deep sequencing also did not guarantee a higher than average score with the *G. polynesiensis* transcriptome analysis resulting in a score of 68.3%. Several of the transcriptomes had very low scores such as the *A. andersonii* (33.3%) and *P. minimum* strain 2233 (39.2%) that correlated to a lower number of assembled contigs (1M and 500k, respectively) compared to those with high scores (1.8M for *A. tamarense*).

### Domain tabulation

The core dinoflagellates were shown to have many more synthase modules relative to the syndinales and outgroups. *Lingulodinium polyedra* (F. Stein) possessed the most domains (total HMM hits) and most domain types (unique functional group hits, *e.g.* "ketosynthase") with 2407 and 1304, respectively (Table 1-1). In total there were 55,818 HMM hits with sufficient scores (<1e-10) across all transcriptomes (*including those with low BUSCO scores*) with a median value of 859 per transcriptome, although around 40% of these were tetratricopeptide repeats predominantly occurring in the core dinoflagellates. When the number of modular synthase hits (again excluding tetratricopeptide repeats) was reduced to functional domains by taking the maximum score across all HMMs for the same domain there were 27,424 domains in the core dinoflagellates compared to 1332 for the outgroup

species, or an average of 669 and 222 per transcriptome, respectively (Table 1-2).

| | Summary of Domain Types | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Species** | **Adenylation** | **Ketosynthase** | **Ketoreductase** | **Dehydratase** | **Enoyl Reductase** | **Thioesterase** | **Thiolation** | **Acyl_Transferase** | **FSH1** | **Condensation** | **GNAT** | **TTPR** | **SUM** | **SUM (no TTPR)** |
| Akashiwo_sanguineum | 238 | 168 | 123 | 10 | 194 | 11 | 74 | 58 | 41 | 5 | 0 | 561 | 1483 | 922 |
| Alexandrium_catenella | 97 | 172 | 74 | 0 | 57 | 13 | 77 | 104 | 36 | 0 | 0 | 767 | 1397 | 630 |
| Alexandrium_margalefi | 129 | 238 | 117 | 10 | 84 | 22 | 86 | 129 | 46 | 1 | 0 | 715 | 1577 | 862 |
| Alexandrium_monilatum | 161 | 163 | 113 | 9 | 62 | 17 | 87 | 79 | 41 | 1 | 0 | 723 | 1456 | 733 |
| Alexandrium_tamarense | 241 | 188 | 165 | 6 | 117 | 23 | 95 | 145 | 69 | 1 | 0 | 785 | 1835 | 1050 |
| Amphidinium_carterae | 109 | 76 | 61 | 27 | 46 | 17 | 61 | 17 | 15 | 10 | 5 | 304 | 748 | 444 |
| Amphidinium_klebsii | 115 | 63 | 74 | 27 | 38 | 25 | 60 | 17 | 14 | 8 | 3 | 320 | 764 | 444 |
| Amphidinium_massartii | 85 | 68 | 67 | 9 | 32 | 8 | 42 | 16 | 15 | 6 | 2 | 285 | 635 | 350 |
| Azadinium_spinosum | 163 | 342 | 148 | 25 | 93 | 29 | 144 | 166 | 84 | 0 | 0 | 1217 | 2411 | 1194 |
| Brandtodinium_nutriculum | 128 | 112 | 68 | 7 | 53 | 5 | 51 | 30 | 9 | 1 | 0 | 299 | 763 | 464 |
| Ceratium_fusus | 277 | 74 | 123 | 9 | 90 | 49 | 33 | 48 | 19 | 2 | 0 | 148 | 872 | 724 |
| Crypthecodinium_cohnii | 226 | 72 | 113 | 27 | 179 | 6 | 52 | 77 | 19 | 3 | 0 | 278 | 1052 | 774 |
| Dinophysis_acuminata | 119 | 248 | 96 | 3 | 70 | 14 | 127 | 108 | 55 | 1 | 0 | 1013 | 1854 | 841 |
| Durinskia_baltica | 150 | 74 | 70 | 6 | 45 | 4 | 39 | 29 | 8 | 0 | 0 | 294 | 719 | 425 |
| Gambierdiscus_excentricus | 101 | 35 | 92 | 4 | 71 | 14 | 64 | 63 | 42 | 0 | 0 | 657 | 1143 | 486 |
| Gambierdiscus_polynesiensis | 125 | 177 | 106 | 12 | 90 | 20 | 106 | 169 | 34 | 0 | 0 | 655 | 1494 | 839 |
| Glenodinium_foliaceum | 237 | 92 | 89 | 4 | 71 | 6 | 49 | 46 | 11 | 2 | 0 | 271 | 878 | 607 |
| Gymnodinium_catenatum | 138 | 36 | 51 | 3 | 46 | 7 | 36 | 21 | 14 | 0 | 0 | 39 | 391 | 352 |
| Gyrodinium_instriatum | 559 | 83 | 157 | 29 | 84 | 12 | 200 | 81 | 57 | 1 | 0 | 359 | 1622 | 1263 |
| Heterocapsa_arctica | 71 | 129 | 66 | 11 | 25 | 6 | 58 | 54 | 9 | 2 | 0 | 288 | 719 | 431 |
| Heterocapsa_rotundata | 66 | 120 | 50 | 6 | 28 | 9 | 45 | 36 | 7 | 4 | 0 | 269 | 640 | 371 |
| Karenia_brevis | 234 | 166 | 163 | 9 | 110 | 9 | 168 | 75 | 71 | 6 | 2 | 601 | 1614 | 1013 |
| Karlodinium_veneficum | 260 | 95 | 117 | 5 | 82 | 4 | 89 | 41 | 23 | 4 | 0 | 436 | 1156 | 720 |
| Kryptoperidinium_foliaceum | 370 | 129 | 120 | 5 | 98 | 14 | 90 | 68 | 21 | 3 | 0 | 489 | 1407 | 918 |
| Lingulodinium_polyedrum | 217 | 371 | 239 | 3 | 97 | 25 | 137 | 210 | 82 | 6 | 0 | 1192 | 2579 | 1387 |
| Noctiluca_scintilans | 85 | 67 | 75 | 6 | 50 | 4 | 26 | 23 | 9 | 0 | 0 | 192 | 537 | 345 |
| Pelagodinium_beii | 145 | 102 | 75 | 6 | 60 | 5 | 69 | 32 | 15 | 0 | 0 | 321 | 830 | 509 |
| Peridinium_aciculiferum | 144 | 79 | 76 | 4 | 56 | 7 | 59 | 34 | 22 | 1 | 0 | 375 | 857 | 482 |
| Prorocentrum_hoffmanianum | 143 | 169 | 92 | 12 | 63 | 8 | 55 | 51 | 31 | 1 | 0 | 705 | 1330 | 625 |
| Prorocentrum_micans | 252 | 64 | 158 | 13 | 91 | 36 | 94 | 33 | 26 | 0 | 0 | 181 | 948 | 767 |
| Prorocentrum_minimum_1329 | 127 | 42 | 95 | 3 | 49 | 16 | 31 | 21 | 18 | 2 | 0 | 167 | 571 | 404 |
| Protoceratium_reticulatum | 147 | 182 | 68 | 3 | 61 | 11 | 78 | 66 | 26 | 2 | 0 | 650 | 1294 | 644 |
| Pyrodinium_bahamense | 133 | 145 | 95 | 6 | 62 | 10 | 67 | 104 | 22 | 0 | 0 | 538 | 1182 | 644 |
| Scrippsiella_hangoei | 202 | 122 | 128 | 14 | 104 | 15 | 120 | 41 | 38 | 1 | 0 | 461 | 1246 | 785 |
| Scrippsiella_hangoei_like | 181 | 92 | 95 | 4 | 63 | 7 | 83 | 37 | 25 | 1 | 0 | 422 | 1010 | 588 |
| Scrippsiella_trochoidea_CCMP3099 | 313 | 141 | 152 | 14 | 105 | 19 | 154 | 63 | 32 | 7 | 0 | 487 | 1487 | 1000 |
| Symbiodinium_sp_B1 | 96 | 53 | 88 | 7 | 46 | 7 | 95 | 17 | 19 | 1 | 0 | 349 | 778 | 429 |
| Symbiodinium_sp_C1 | 125 | 70 | 91 | 7 | 58 | 7 | 65 | 25 | 18 | 1 | 0 | 362 | 829 | 467 |
| Symbiodinium_sp_CCMP421 | 187 | 147 | 101 | 13 | 81 | 13 | 117 | 49 | 23 | 0 | 0 | 586 | 1317 | 731 |
| Symbiodinium_sp_cladeA | 78 | 74 | 54 | 24 | 33 | 6 | 33 | 21 | 8 | 7 | 1 | 124 | 463 | 339 |
| Symbiodinium_sp_Mp | 120 | 66 | 76 | 2 | 42 | 9 | 62 | 23 | 21 | 0 | 0 | 400 | 821 | 421 |
| **SUM (core dinoflagellates)** | 7094 | 5106 | 4181 | 404 | 2986 | 549 | 3278 | 2527 | 1195 | 91 | 13 | 19285 | 46709 | 27424 |
| Chromera_velia | 24 | 1 | 20 | 0 | 22 | 1 | 1 | 1 | 2 | 0 | 0 | 13 | 85 | 72 |
| Hematodinium_sp | 119 | 61 | 91 | 18 | 106 | 1 | 52 | 25 | 27 | 3 | 0 | 8 | 511 | 503 |
| Oxyrrhis_marina | 129 | 21 | 71 | 0 | 60 | 3 | 11 | 21 | 9 | 0 | 0 | 56 | 381 | 325 |
| Oxyrrhis_marina_LB1974 | 105 | 25 | 48 | 0 | 43 | 1 | 14 | 24 | 6 | 0 | 0 | 58 | 324 | 266 |
| Perkinsus_marinus | 16 | 0 | 9 | 13 | 15 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 55 | 55 |
| Triceratium_dubium | 46 | 9 | 28 | 0 | 25 | 0 | 0 | 1 | 1 | 1 | 0 | 165 | 276 | 111 |
| **SUM (outgroups)** | 439 | 117 | 267 | 31 | 271 | 6 | 78 | 72 | 47 | 4 | 0 | 300 | 1632 | 1332 |
| **TOTAL** | 7533 | 5223 | 4448 | 435 | 3257 | 555 | 3356 | 2599 | 1242 | 95 | 13 | 19585 | 48341 | 28756 |
| **Values shown represent the count of each domain type in each transcriptome or the sum when designated.** | | | | | | | | | | | | | | |
| Abbreviations: FSH1 = fision yeash serine hydrolase 1, GNAT = GCN5-associated N-acetyl transferase , TTPR = tetratricopeptide repeat | | | | | | | | | | | | | | |

When the two dinoflagellate outgroup species *Hematodinium* and *Oxyrrhis* are removed and the remaining outgroup alveolates are taken separately: the average drops even further to 79. The largest difference was in the thioesterase domains that were thirteen times more abundant in the core dinoflagellates and often only a single copy in the outgroup species. Thiolation and ketosynthase domains were also much more abundant in the core dinoflagellates with a more than six-fold increase indicating that core dinoflagellates possess a higher synthetic capacity than other dinoflagellates and alveolates on average.

The GNAT and condensation domains from the TeCATe transcript were poorly represented in the core dinoflagellates and absent from the outgroup species (Table 1-2). This is likely due to the low number of BLAST results (five for the GNAT, six for condensation domain 1, and nine for condensation domain 2) from the primary search in *A. carterae* that limited the creation of a robust HMM for GNAT and condensation domains. By contrast the adenylation domain from the TeCATe transcript resulted in 41 BLAST hits. The HMM search was still more sensitive than BLAST alone with a total of 13 transcripts detected with a GNAT domain using the HMM with an e-value cutoff of 1e-10 versus 8 using BLAST with no cutoff among all transcriptomes (data not shown). Still, these domains may have been under-sampled especially for taxa more distantly related to *A. carterae*. The GNAT and condensation domains were the least represented among all other domains with

adenylation and ketosynthase having the highest relative abundance across taxonomic groups (Figure 1-2). Thiolation, acyltransferase, enoyl reductase, and ketoreductase

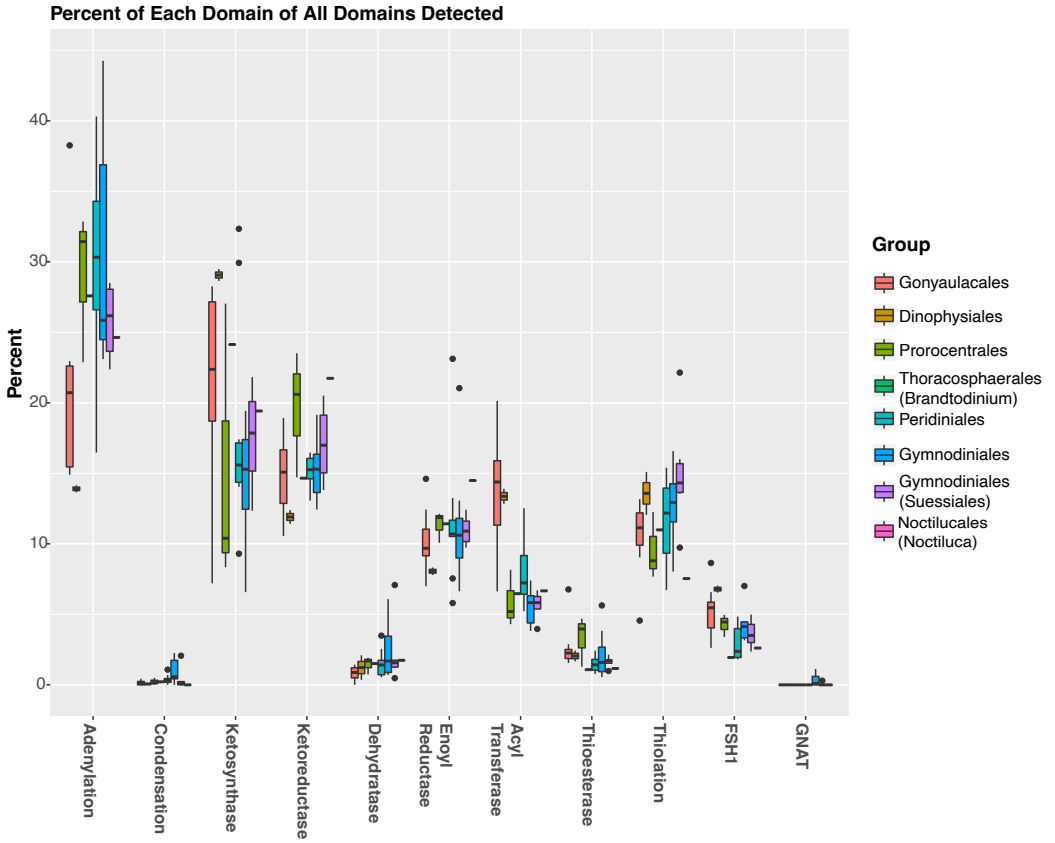



**Figure 1-2: The percent of each domain relative to all domains detected in dinoflagelates using the dinoflagellate HMMs.**
The relative dinoflagelate domain abundance for each domain is shown with the percent shown on the Y-axis and boxplots of the values when more than one species was present in each group with black circles denoting outlier values. Dinoflagellates were grouped taxonomically by their order and colored according to the legend on the right. The domains are shown on the X-axis excluding the tetratricopeptide repeat domains that were used in the calculation but were frequently not associated with any of the modular synthase domains.

domains were also usually well represented while dehydratase, thioesterase, and the FSH1 serine hydrolase were in relatively low abundance.

This picture changes when looking at multi-domain transcripts (transcripts with more than one domain type, not including multiple domain hits of the same domain type), where roughly a third of dehydratase domains and half of thiolation domains were found in these multi-domain transcripts while adenylation, ketosynthase, ketoreductase, and enoyl reductase domains are predominantly found as single domains (Figure 1-3). These trends frequently held across taxonomic groupings except

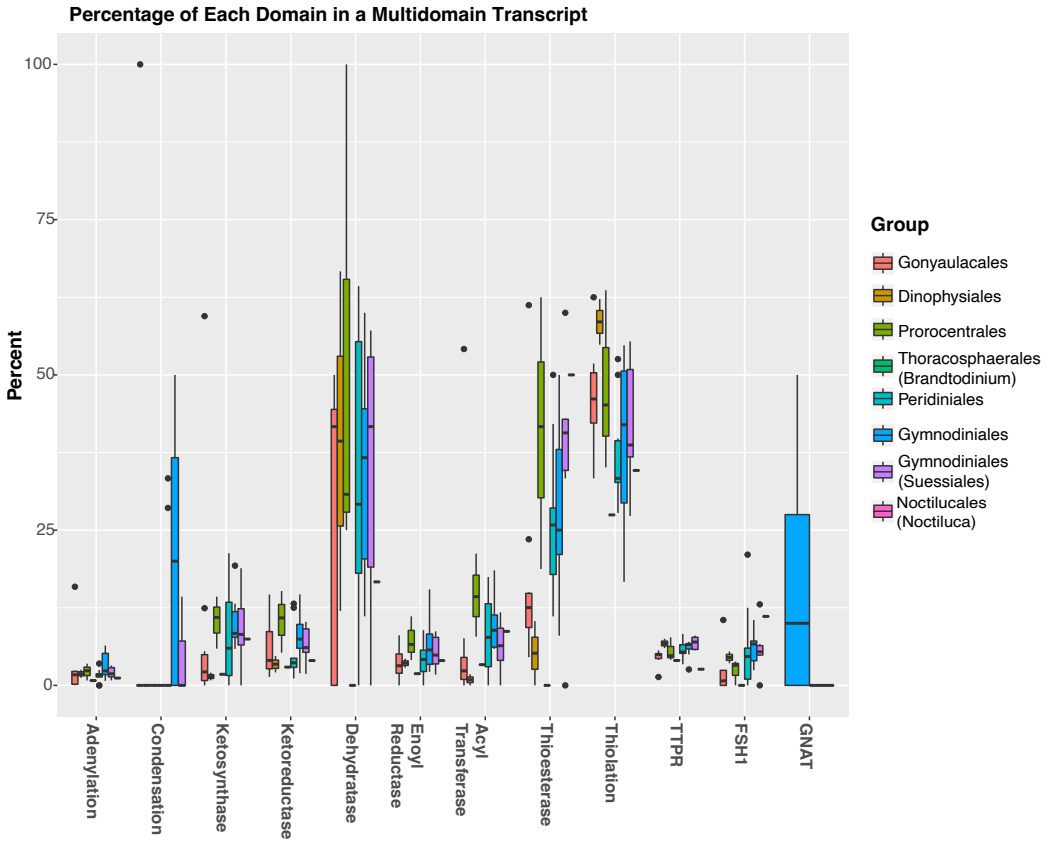



**Figure 1-3: The percent of each domain in a multidomain transcript relative to the total number of each domain found in dinoflagellates.** The relative abundance of each domain type out of the total number of that domain (excluding tandemly repeated domains wihout other domain types) is shown as a percent on the Y-axis and a box plot wheh multiple species are present with black circles denoting outlier values. Dinoflagellates were grouped taxonomically by their order and colored according to the legend on the right. The GNAT domain was only present in two taxonomic orders, the Gymnodiniales and the Suessiales, and the boxplots are thus drawn with a different width.

for thioesterases, which were found 10-15% of the time in multi-domain transcripts for the Gonyaulacales and Dinophysiales and a quarter to a third of the time as multi-domain transcripts in the other taxonomic groups. Multi-domain transcripts were the exception in core dinoflagellates accounting for 8.34% of all domain types (excluding tetratricopeptide repeats) with an average of 13.84% for each species and domain type combination.

The relative abundance of modular synthase domains was similar across species with no obvious differences in documented toxin-producing species. The principal components plot based on domain counts and colored by toxin type was used to demonstrate this (Figure 1-4A) and showed a general clustering of all species,

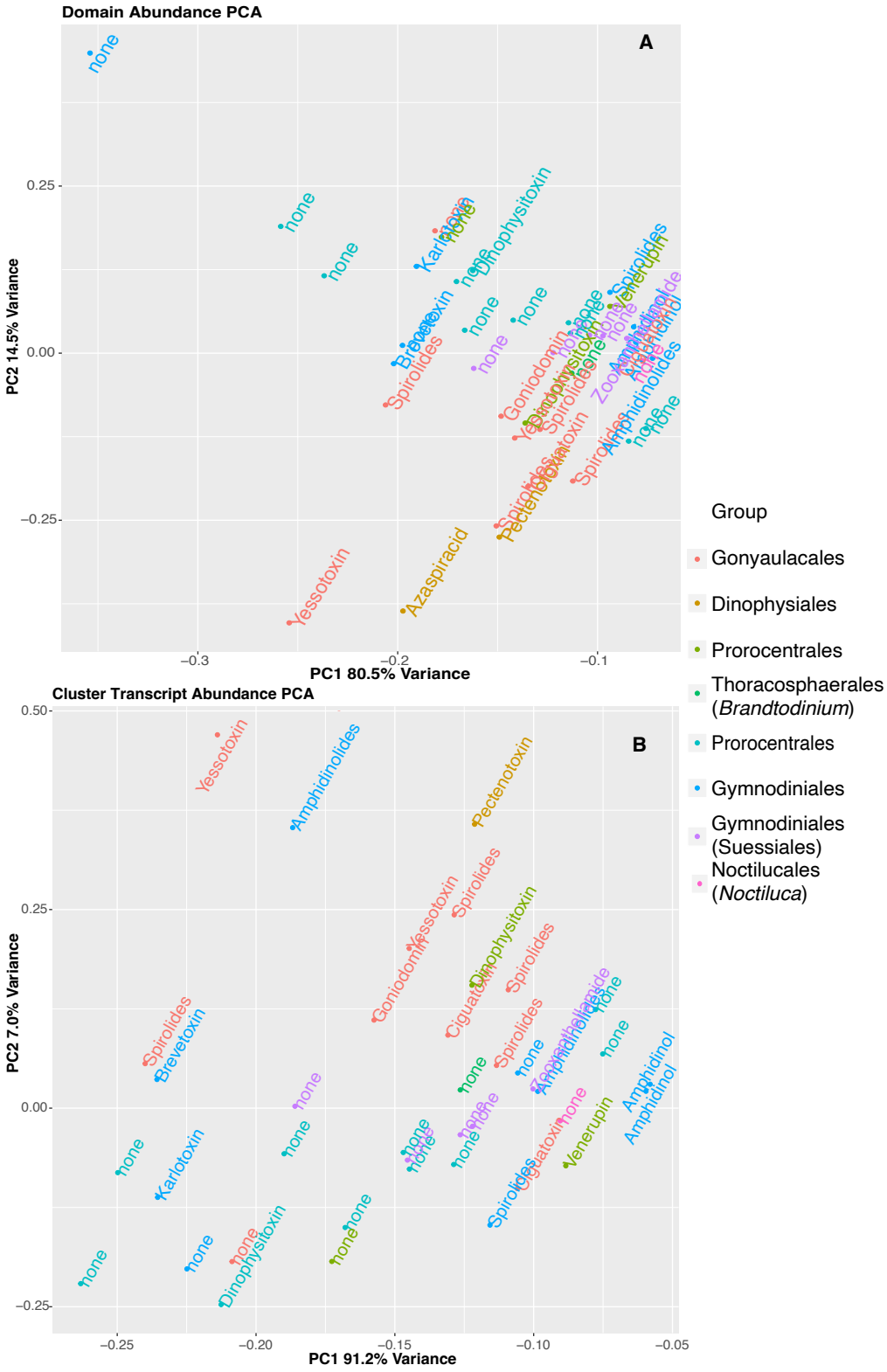



**Figure 1-4: Principal component plots of the overall domain counts within dinoflagellates (A) and the breakdown of those counts within clusters of highly similar sequences (B).**
Principal components are shown for the total domain counts among taxonomic groups of dinoflagellates (A) as well as the count of individual transcript within clusters of very similar domain sequences (B). Principal component 1 is shown on the X-axis and component 2 is shown on the Y-axis with the relative contribution of each component shown next to each axis. The individual points represent an individual transcriptome that is colored for the order level taxonomy of each species from which the transcriptome was samples shown on the legend on the right. Each point is also labeled with a widely recognized toxin that is made by that secies.

irrespective of toxin type except for three species: *Gyrodinium instriatum*, that does not make a known toxin and has a higher proportional number of adenylation domains (Axis 1 outlier on the far left), and *Lingulodinium polyedra* and *Azadinium spinosum* that make yessotoxin and azaspiracids, respectively, and have a proportionally higher number of ketosynthase domains (Axis 2 outliers on the bottom). There were, however, lineage specific differences for specific domain types. Thiolation domains were more relatively abundant in the more basal Gymnodiniales compared to acyl transferase domains in a decreasing trend to the more distal Gonyaulacales (Figure 1-5). This is also visible in the plot of domains as percentages.

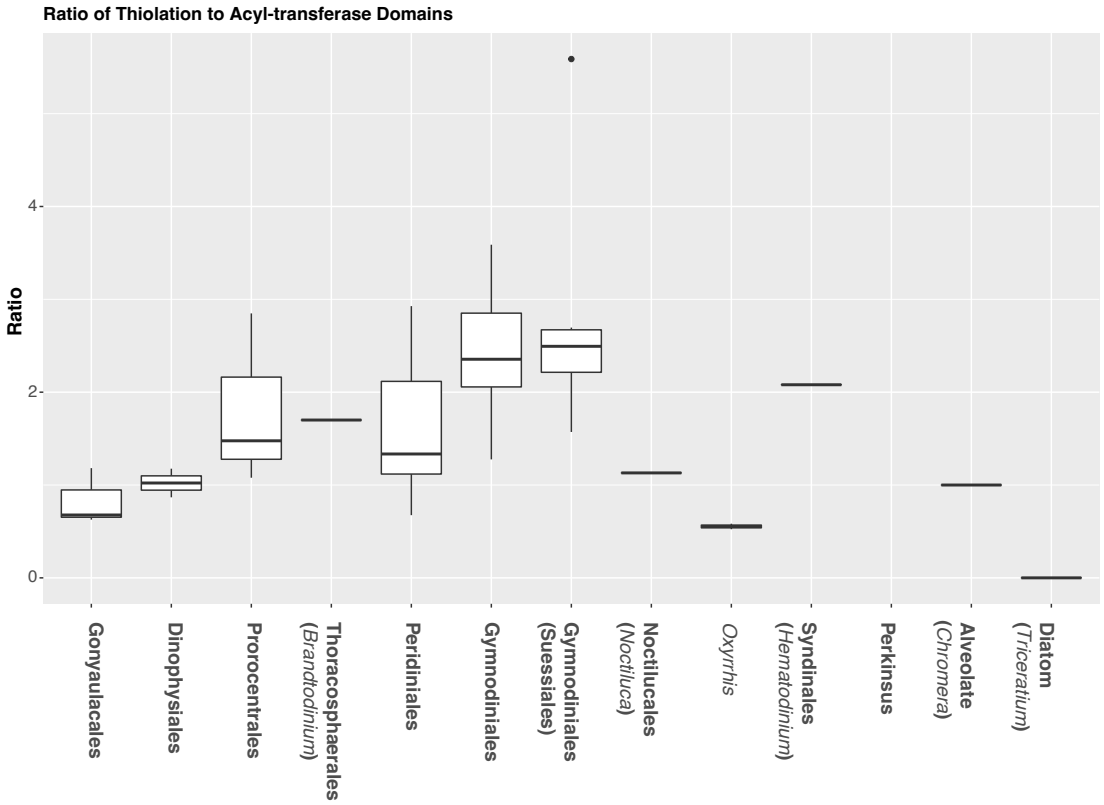



**Figure 1-5: The ratio of thiolation to acyl-transferase domains in dinoflagellate and outgroup taxa.**
The relative abundance of thiolation and acyl-transferase domains are shown as a boxplot with the ration of thiolation to acyl-transferases on the Y-axis. The X-axis shows the taxonomic grouping with dinoflagellates grouped by Order and single species given as their genus in parentheses with the exception of *Oxyrrhis*. The Syndiniales and other outgroup taxa are also shown represented by a single individual.

The acyltransferase domains make up a much higher percentage in the Gonyaulacales and Dinophysiales versus other taxonomic groups (Figure 1-2), although this is less obvious for the thiolation domains. There is also a high average number of thiolation domains in a transcript when comparing the Gymnodiniales to the other taxonomic groups (Figure 1-6).

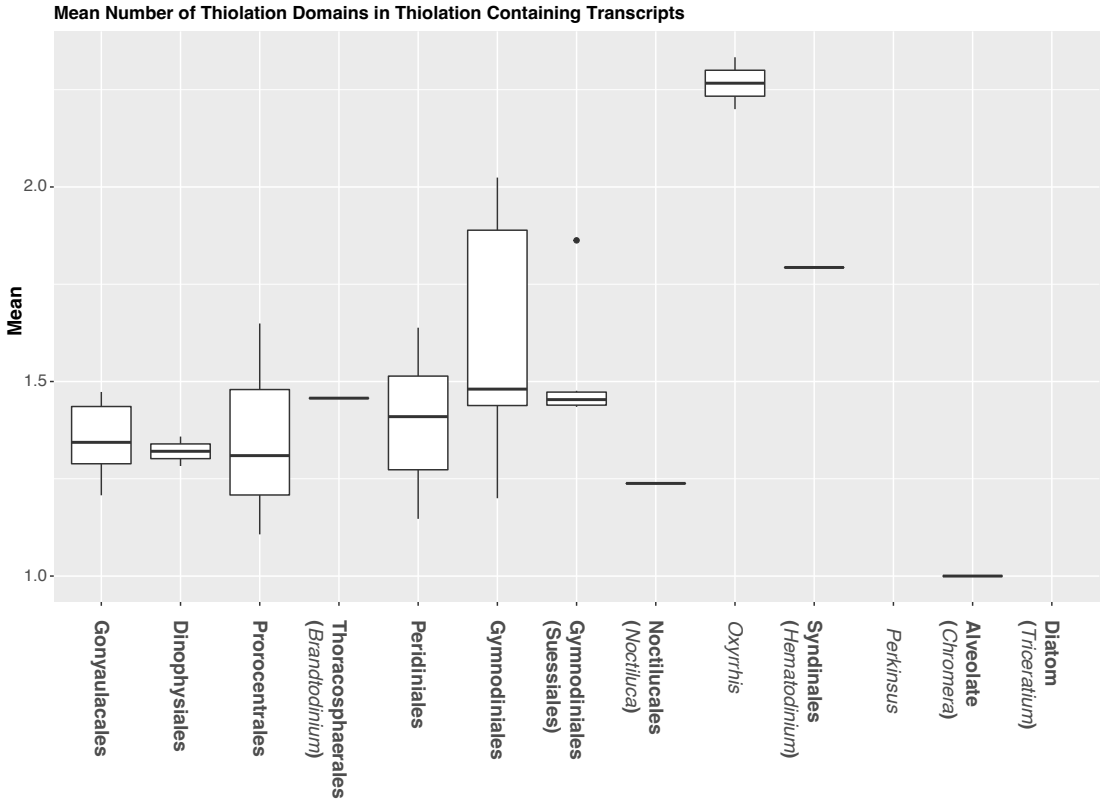



**Figure 1-6: The mean number of thiolation domains in all thiolation domain containing transcripts.**
The average number of thiolation domains per transcript in all thiolation domain containing transcripts is shown on the Y-axis while the X-axis shows the taxonomic grouping with dinoflagellates grouped by Order and single species given as their genus in parentheses with the exception of *Oxyrrhis*. Transcripts did not have to have any other domain type in order to be counted and many transcripts contained thiolation domains exclusively.

Although tetratricopeptide repeats were found in almost every transcriptome, the abundance was much higher in the core dinoflagellates (19,285 in core dinoflagellates versus 300 in outgroups or a four-fold increase on average per transcriptome) and the combination of this repeat in transcripts with thiolation domains was only found in the core dinoflagellates (Figure 1-7). The number of repeats varied within a

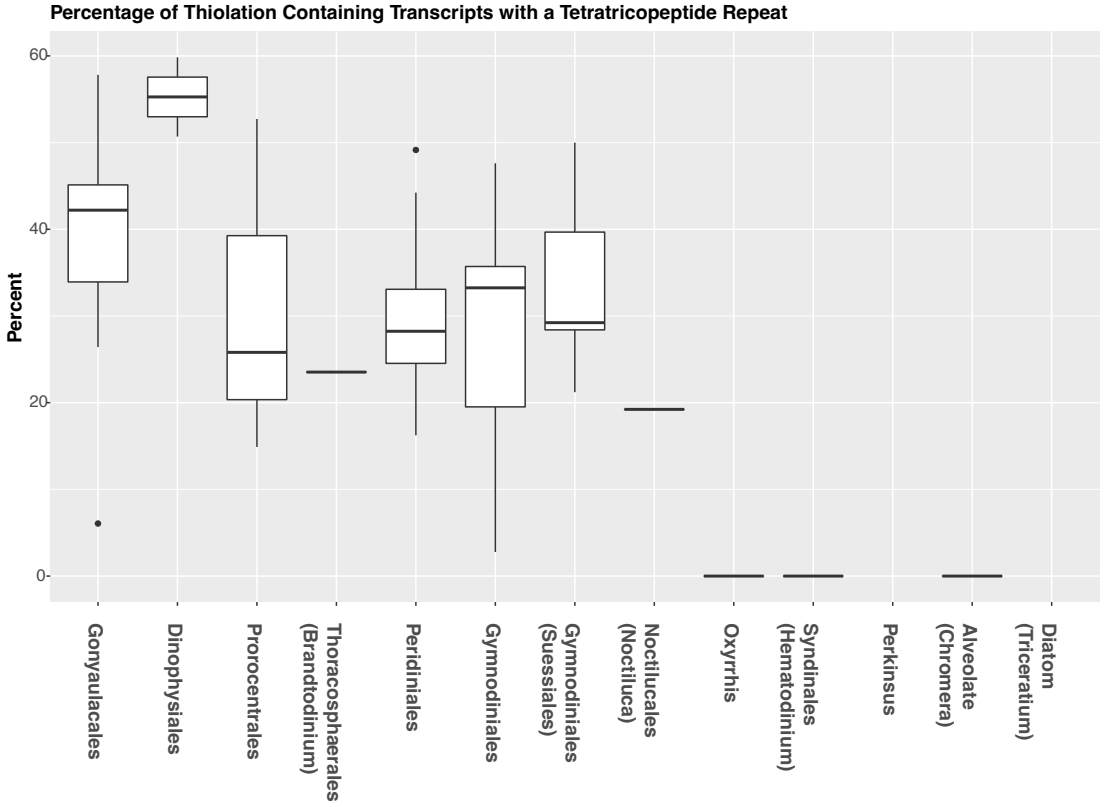



**Figure 1-7: Percentage of transcripts containing a thiolation domain and tetratricopeptide repeats in dinoflagellates and outgroup species.**
The percentage of thiolation domain conatining transcripts that also contain a tetratricopeptide repeat are shown with the percentage on the Y-axis as boxplots when more than one species is present. The X-axis shows the order level taxonomy of the dinoflagelate species whith the exception of *Oxyrrhis* and *Perkinsus* where the phylogenetic placement is less certain. The remaining outgroup species are are described as their common phylum name followed by the genus level taxonomy in parentheses.

transcript from one to twenty and the repeat number distribution is approximately log normal in shape with low repeat numbers being very frequent (Figure 1-8A).

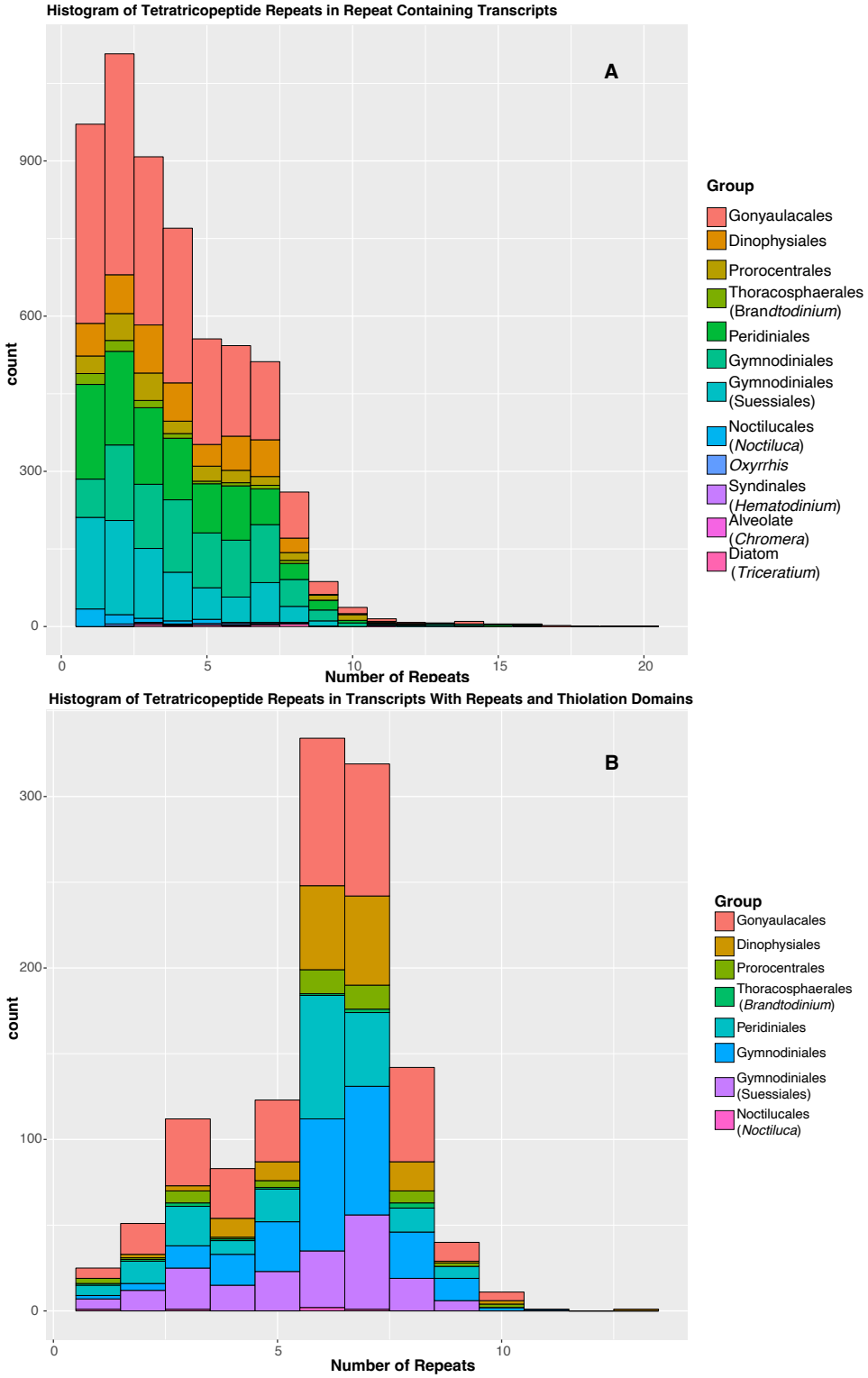



**Figure 1-8: Histogram of the number of tetratricopeptide repeats in repeat containing transcripts.**
The count of the tetratricopeptide containing transcripts is shown on the Y-axis while the number of repeats in those transcripts is shown on the X- axis. The upper panel (A) shows the histogram of tetratricopeptide repeats in all repeat containing transcripts while the lower panel (B) shows the histogram for transcripts that also contains a thiolation domain. This combination of tetratricopeptide repeats was only observed in core dinoflagellates and thus the legends for the two panels are not identical. The upper legend for panel A includes dinoflagellates grouped by order with individual specimens given as their genus in parentheses along with outgroup species while the lower legend for panel B only includes core dinoflagellates.

This distribution changes dramatically when looking at repeat numbers in transcripts with a thiolation domain where six and seven member repeats are very frequent approximating a t-distribution (Figure 1-8B). The relative number of repeats among each taxonomic group did not vary greatly and approximated the relative total number of domains found.

### Domain Clustering and Gene Duplication

The protein sequence clustering provides a three-dimensional relationship where more similar sequences are more closely spaced. If the points are close enough to pass a calculated probability threshold of $p < 0.001$ then a line is drawn denoting significant similarity and a group of points with interconnecting lines was identified as a cluster. While these data do not reflect inferred ancestry of the sequences like a phylogenetic tree would, the relationships are not forced into a bifurcating arrangement. This is helpful in visualizing many, very similar sequences such as dinoflagellate domains where there is an abundance of gene duplication and strict orthology is difficult to ascertain.

The domains retrieved by HMM searches were compared based on the number of clusters, where a large number of clusters implies a high degree of inferred functional diversity, and the size of the clusters is an indication of the amount of gene duplication for that function. Clusters were also searched for annotated multi-domain transcripts to compare clusters across and within each domain. The number of sequences used for clustering varied substantially between domain type with 15,865 adenylation; 10,118 ketoreductase; 9832 ketosynthase; 7854 thiolation; 7025 enoyl-reductase; 3324 dehydratase; 2492 acyl transferase; and 1085 thioesterase domains, following dereplication of sequences. There are also likely some false positives from the HMM search with 7887 of the 202,024 total sequences containing internal stop codons that may be from the translation of a spurious open reading frame coincidently similar to the HMM. These false positives as well as some truncated sequences appear in the clusterings as outlying spots. Also, sequencing depth may artificially inflate or deflate the size of each cluster. The BUSCO scores for each transcriptome used in clustering were similar so this is not likely to be a dramatic effect. Likewise, this is unlikely to affect the number of clusters, only the size, since only a relatively large number of similar sequences would generate a cluster.

The thiolation domain clustering represents an example of low diversity and low copy number with only two clusters formed when the acyl carrier protein was added into the dataset and a small trail off of the largest cluster containing one of the ZmaK thiolation domains (Figure 1-9). The main cluster has several subclusters, one

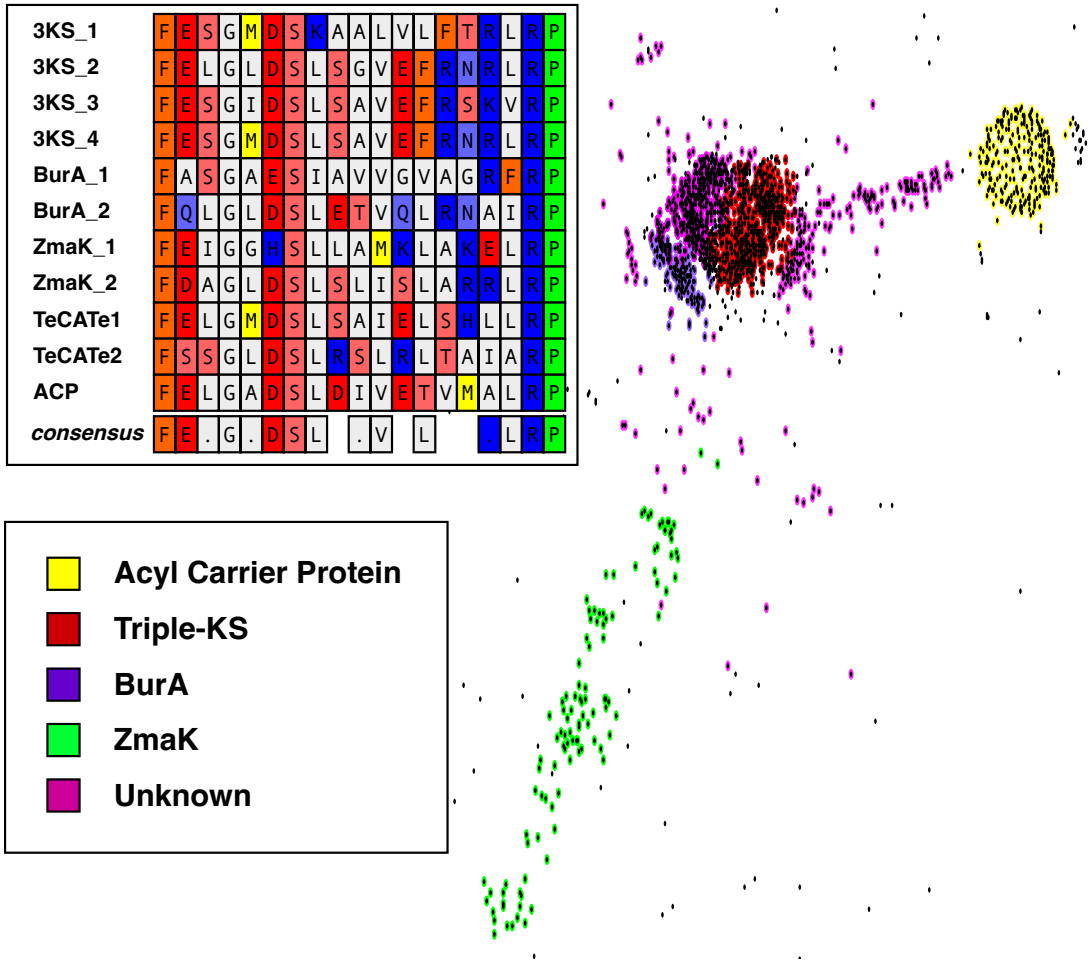



**Figure 1-9. Cluster plot of the thiolation domains.**
A clustering of the protein sequences for dinoflgellate thiolation domains is shown. The acyl carrier protein was added back into the analysis since this was not recovered in the hidden Markov model search and is colored yellow. Thiolation domain clusters containing the Triple-KS and ZmaK_2 (red), BurA (purple), and other unidentified thiolation domains form the central cluster while a cluster containing the ZmaK_1 thiolation domain (green) is shown extending down and to the left of the central cluster. An alignment of the reference thiolation domains from *Amphidinium carterae* is shown in the upper left.

containing BurA transcripts, a second containing the triple KS transcripts, a third that has several transcripts with adenylation and ketosynthase domains that are not one of the four annotated transcripts from Figure 1-1. There is some resolution of the subclusters separating the annotated transcripts from each other, but they are tightly linked with many internal edges. The other main cluster exclusively contains acyl carrier protein sequences from each of the transcriptomes. These differences can be seen when viewing an alignment of the binding sites from *A. carterae* for the phosphopantetheinyl transferase that activates the thiolation domain (Figure 1-9 insert). Most of the domains have similar positively charged residues following and negatively charged residues preceding the invariant serine that serves as the

phosphopantetheinate attachment site. For one of the ZmaK sites, the negatively charged residue is instead positively charged and for the acyl carrier protein there is a methionine. The acyl carrier protein from *Escherichia coli* is also shown with a methionine to show how conserved this residue is in the acyl carrier protein making this gene easy to distinguish from other modular synthases. Thus, there is a clear segregation of fat synthesis from other small molecule synthesis in the thiolation domain clusters irrespective of any implied gene origin such as horizontal gene transfer from bacteria in the case of BurA and ZmaK. This can also validate the data to some degree as the ACP cluster contains approximately 230 sequences averaging four per transcriptome. *Amphidinium carterae* has three readily identifiable acyl carrier proteins agreeing with the expected number of sequences in the ACP cluster.

Other domains with limited clusters include the dehydratase, enoyl reductase, and thioesterase domains (Figure 1-10). The dehydratase domains (Figure 1-10A) form a single cluster of sequences including those from the triple-KS and ZmaK transcripts as well as an ancillary cluster that is annotated as a "Domain of Unknown Function" by the NCBI COG database and is similar to dehydratases involved in tyrosine metabolism by BLAST. The low copy number of the dehydratase domain is in contrast to the numerous enoyl reductase clusters (Figure 1-10B). Diversity is still low with a single cluster containing the ZmaK transcript and two other clusters that do not contain annotated transcripts but there are also many sequence fragments that form satellite points and do not cluster. The thioesterase domain clustering (Figure 1-10C)

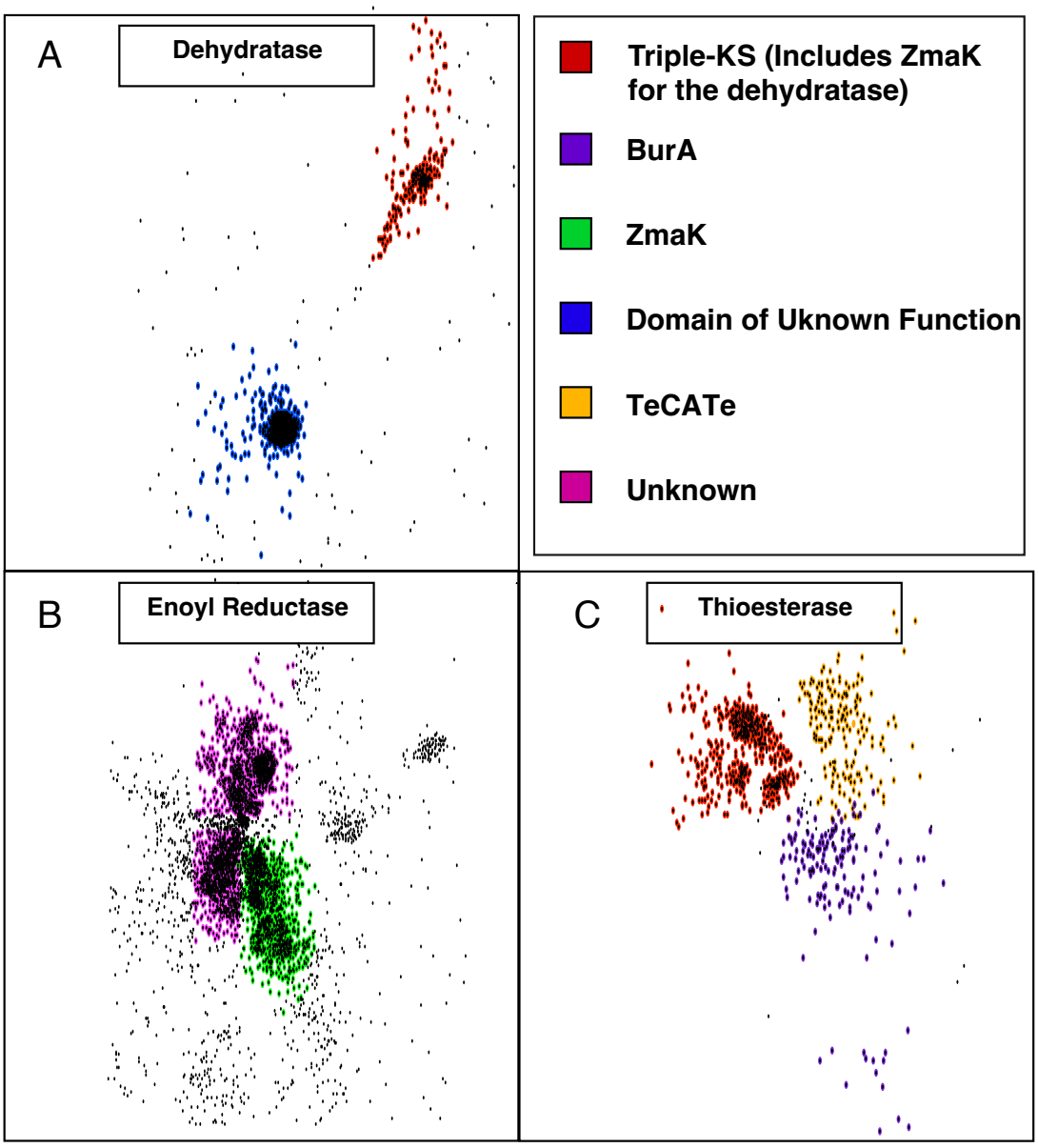



**Figure 1-10. Protein sequence clusters of dehydratase, enoyl reductase, and thioesterase domains.**
The clusterings of the dehydratase (A), enoyl reductase (B), and Thioesterase (C) protein sequences are shown. All domains exhibit relatively low levels of duplication. The dehydratase domain has two clusters including one of "unknown function" that is similar to amino acid dehydratases according to the NCBI ortholog database. The enoyl reductase domain has three clusters including ZmaK and two of unknown similarity The triple KS transcript is found in several different clusters depending on the species. The simplest domain is thioesterase with clusters for the BurA, triple KS, and TeCATe transcripts from *Amphidinium carterae*.

contains three low abundance clusters all containing annotated transcripts.

For the ketosynthase and adenylation domains the domain count was very high with several unknown clusters (Figure 1-11), similar to the enoyl reductase

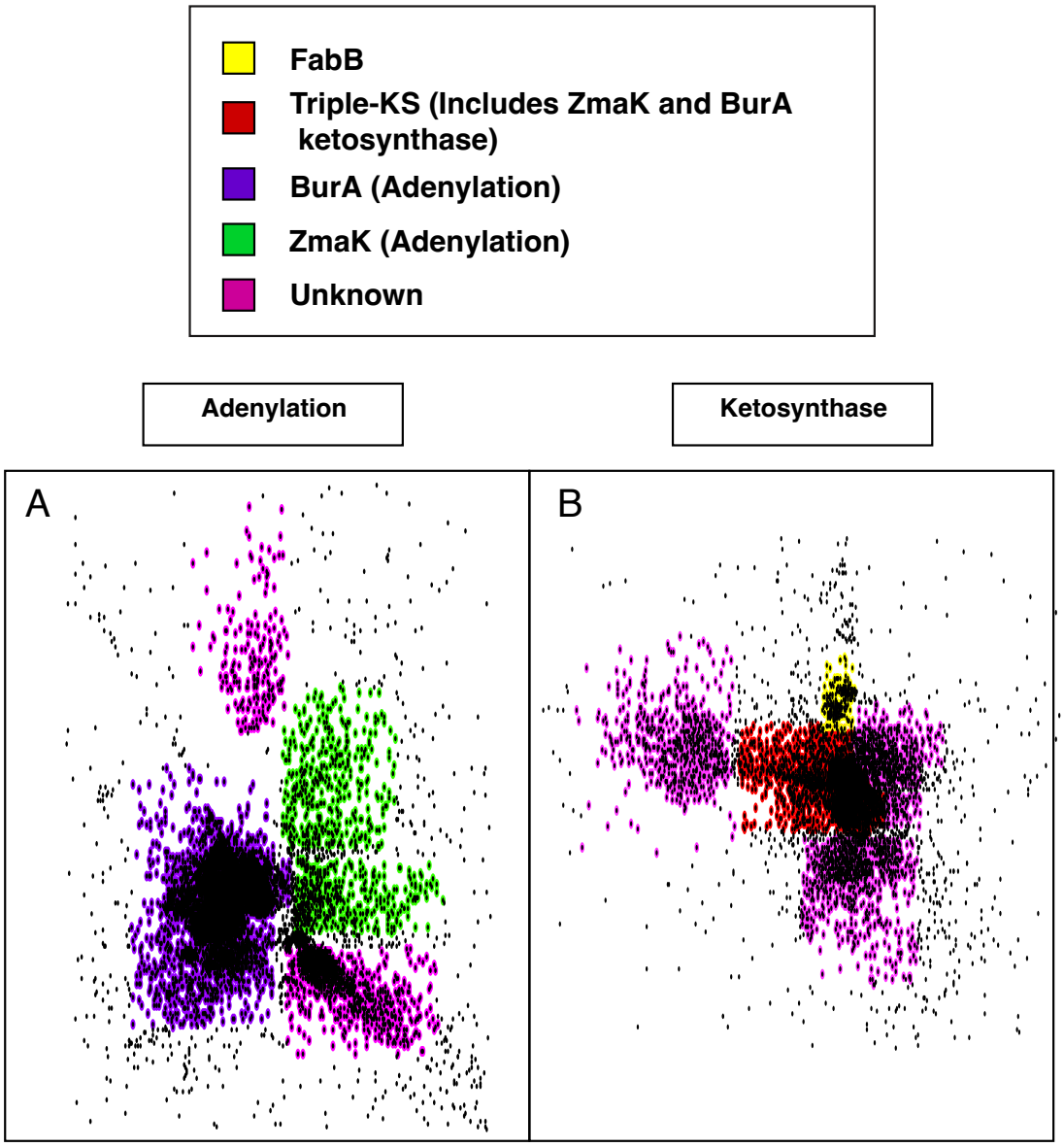



**Figure 1-11. Adenylation and Ketosynthase protein sequence clusters.**
The clusterings of the adenylation (A) and ketosynthase (B) protein sequences are shown. Both exhibit gene duplication of domains from the triple KS, BurA, and ZmaK transcripts forming two large clusters of adenylation domains and one very dense cluster of ketosynthase domains. Several clusters of unknown similarity are also apparent as well as the FabB ketosynthase cluster that is involved in lipid synthesis and does not appear to be heavily duplicated.

clustering but with a larger number of clusters. For the adenylation domain clustering, (Figure 1-11A) the number of sequences was the largest with over fifteen-thousand unique sequences. The adenylation domains from the ZmaK and BurA transcripts were found in separate clusters but for the ketosynthase domain (Figure 1-11B) the annotated transcripts all occupy a single cluster with poor resolution of subclusters

and include the condensation domains from the TeCATe transcript. Both domains produced several clusters that do not contain annotated transcripts and the ketosynthase domains involved in fat synthesis labeled “FabB” form a distinct low abundance cluster.

This pattern of large clusters of single domain transcripts that are similar to domains from the annotated transcripts and very small conserved clusters of fat synthesis genes appears to reverse for the acyl transferases and ketoreductases (Figure 1-12). Despite acyl transferases being one of the lowest abundance domains and

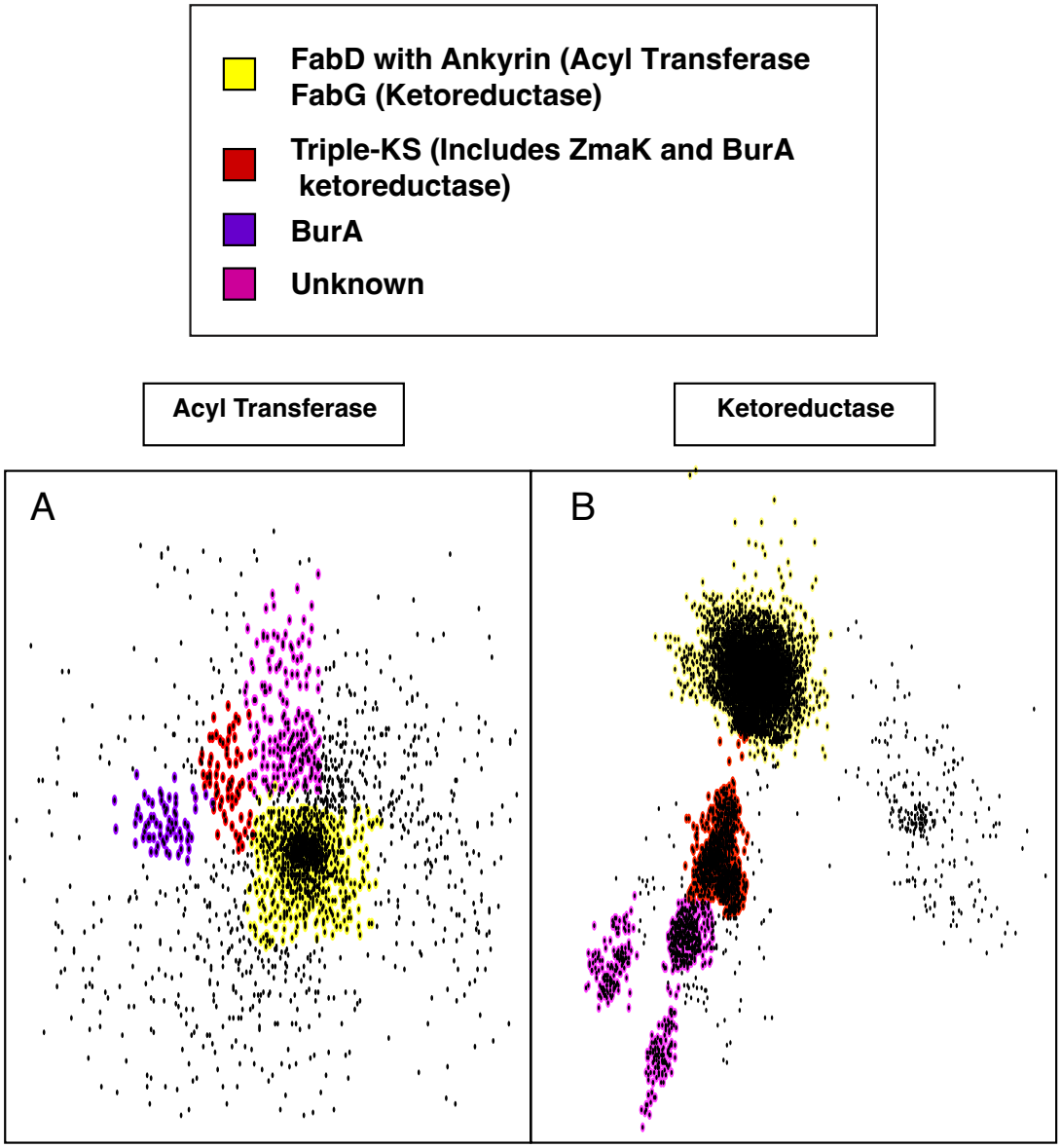



**Figure 1-12. Acyl transferase and ketoreductase protein sequence clusters.**
The clusterings of the acyl transferase (A) and ketoreductase (B) protein sequences are shown. Both exhibit gene duplication of apparent genes involved in fat synthesis (yellow) with FabD for the acyl transferases (often with ankyrin domains) and the FabG gene for ketoreductases. Other clusters include the Triple-KS genes (also including the ZmaK and BurA ketoreductases), a separate BurA acyl transferase cluster, and several clusters of unknown similarity.

ketoreductases one of the highest, both clusterings contain large clusters of sequences similar to the fat synthesis genes FabD (acyl transferase) and FabG (ketoreductase). There are small clusters of acyl transferase domains from the triple KS and BurA transcripts with very few single domain transcripts. The cluster containing FabD like

transcripts is quite large with many of these transcripts containing ankyrin repeats that promote and regulate protein-protein interactions (Mosavi, Cammett, Desrosiers, & Peng, 2004). There is also a small cluster of single domain transcripts that are not similar to the annotated transcripts. Likewise, the large ketoreductase clusters (Figure 1-12B) demonstrate the annotated transcripts in a single cluster, three clusters that do not contain annotated transcripts, and a final very large cluster containing the FabG gene.

Principal components plots of the counts within each cluster (Figure 1-3B) gave similar results to the overall domain counts with no association between toxins produced or phylogenetic group to principal component positions. Principal component axis 1, which accounted for 91.2% of the variance differed mainly in the expansion of BurA-like domains with species on the left portion of the graph possessing large numbers of BurA-like domains while those on the right had very few. This did not correlate to intact BurA transcripts and the level of expansion was not always consistent, *e.g.* that *Karenia brevis* was found to have 512 BurA-like adenylation domains but only 13 BurA-like thioesterase domains (Table S1-2). In both cases this count was much higher than other species, but not equivalent.

Phylogenetic inference was attempted on the two smallest datasets, acyl transferases and thioesterases, to determine ancestry and compare the results to the clustering results. The resultant trees (Supplementary files 1-1 and 1-2) had zero or near zero bootstrap support for all major and minor branches up to the final bifurcations indicating that determination of ancestry was not possible for these datasets and methods. The highest scoring trees were able to replicate the clustering results to some degree with major clades mirroring the clusters formed. Also, very similar sequences or assembly variants were visible with high bootstrap support at the distal branches.

## Discussion

### Modular Synthases are Abundant in the Core Dinoflagellates

The goal of this study was to investigate the abundance, diversity, domain arrangement, and evolution of enzymes likely to participate in the synthesis of dinoflagellate toxins by focusing on dinoflagellates and the synthetic domains found consistently within dinoflagellates independent of existing model frameworks. There is an inherent need to study these synthetic pathways since dinoflagellate toxins are the largest known natural products with high potential toxicity (Fukatsu et al., 2007; Sasaki et al., 1996), are synthesized using many non-canonical and interesting chemistries (Van Wagoner et al., 2014; Wright et al., 1996), and have potential therapeutic uses (Fukatsu et al., 2007; Javed et al., 2011), but we have very little understanding of how they are synthesized. Based on isotopic labeling studies they are predominantly made of acetate units incorporated by polyketide synthases (Lee et al., 1989; Wright et al., 1996) with the occasional amino acid or other carboxylic acid used as starter and extender units via non-ribosomal peptide synthases (Rasmussen et

al., 2017). In both cases, a condensation reaction incorporates a chemical unit into a growing molecule that is subsequently modified either during elongation or following the synthesis of a large portion of the molecule (Khosla, 2009). To facilitate discussion polyketide synthases and non-ribosomal peptide synthases were combined into the term "modular synthases" to encompass both condensation reactions and the general similarities of their chemistry and genetics. For polyketide synthases, the condensation reaction is performed by a ketosynthase that usually incorporates acetate from malonyl CoA but can also facilitate the addition of other short carbohydrates (Moore & Hertweck, 2002) with the release of carbon dioxide. Non-ribosomal peptide synthases use adenylation domains to pass a specific substrate, often an amino acid, to the condensation domain with microcystin being a common example (Fewer et al., 2007). Adenylation domains can also be found in the same module as ketosynthases in a hybrid system as is the case for BurA and ZmaK that participate in the synthesis of burkholderic acid in *Burkholderia* species and zwittermicin in *Bacillus* species, respectively (Franke et al., 2012; Kevany et al., 2009). The BurA and ZmaK synthetic pathways are also important to mention because the module has been fragmented in their respective bacterial genomes with separate modules occurring on distant regions of the chromosome while BurA serves an unusual role in bridging these pathways. Due to the processive nature of these modular synthases the pathway is generally encoded as syntenic modules and their domains in a more or less linear fashion that can be used to predict the final product of synthesis (Khosla et al., 2009). Although many domains can come into play in a trans fashion (Hertweck et al., 2007), the most common trans-acting domains are acyl transferases and thioesterases. These domains also do not need to be collinear with other synthetic modules and have been shown to be synthetically active when whole genomic sections have been cloned (Piel, 2002). A cursory BLAST analysis of the published *Polarella glacialis* genome (Stephens et al., 2020) show that domains are commonly found in tandem repeats of the same domain with different domains found on different scaffolds with the exception of common multi-domain transcripts such as the triple KS (Table S1-4).

Dinoflagellates regulate their gene expression largely post-transcriptionally (Lidie et al., 2005; Moore & Hertweck, 2002) making linear encoding of the modular synthase domains obsolete. Unsurprisingly, the vast majority of modular synthase domains have been fragmented and duplicated, similar to what has been shown for other gene families in dinoflagellates such as actin and translation initiation factors (Bachvaroff & Place, 2008; Jones et al., 2015). In this study adenylation, ketosynthase, and ketoreductase domains were frequently observed as single domain transcripts, as has been observed previously (Beedessee et al., 2019; Monroe et al., 2010; Van Dolah et al., 2017). Both single and multi-domain transcripts of modular synthases occurred in high abundance in all core dinoflagellates and their distribution was not correlated with taxonomy, toxicity or toxin type (Figure 1-4). This apparently ubiquitous synthetic capacity argues that secondary metabolite synthesis is a common feature of all core dinoflagellates, a theory supported by observations that polyketide synthesis genes are found in species that do not produce known polyketide toxins (Snyder et al., 2003). Similarly, the only phosphopantetheinyl transferase, the enzyme required to activate thiolation domains and initiate secondary metabolite synthesis,

found in all core dinoflagellates was able to activate a NRPS based reporter system indicative of natural product rather than lipid synthesis (Williams, Bachvaroff, & Place, 2020). This is in contrast to syndinian dinoflagellates and other alveolates that had a much lower abundance of synthetic domains with all domains in similar abundance (Table 1-2). This is likely due to serial duplication that is a hallmark of core dinoflagellate evolution (Shoguchi et al., 2013) and has been shown to affect the evolution of the synthetic pathway for saxitoxin in particular (Murray, Diwan, Orr, Kohli, & John, 2015).

### Single Domain Transcripts Exhibit Domain Specific Patterns of Duplication

Multidomain transcripts were observed in all core dinoflagellates. The triple KS and BurA-like transcripts are more or less intact across almost all of the core dinoflagellates and can be readily found by simple domain counting (Table S1-3) or by looking for large transcripts with ketosynthase domains. The ZmaK-like and TeCATe transcripts are less robust, often truncated or have missing domains but are still readily recognizable. The BurA-like and ZmaK transcripts were horizontally transferred from bacteria since they are largely absent in the syndiniales but are present in a number of bacterial species as part of conserved synthetic pathways. Although modular synthases were almost entirely absent from *Amoebophyra* species, multi-domain polyketide synthases were found in *Hematodinium* in this study (Table 1-2) as well as another separate transcriptomics study that determined them to be cytosolic in nature (Gornik et al., 2015). The sequence arrangement is very similar between the *A. carterae* transcriptome and *Hematodinium* genome polyketide synthases. Similarly in *Toxoplasma* and *Cryptosporidium* there are multi-module PKS genes similar to the dinoflagellate triple KS used as a model here that are theorized to process fatty acids (Mazumdar & Striepen, 2007; Zhu et al., 2004). Dinoflagellates are known to make many poly-unsaturated fatty acids (Leblond, Evans, & Chapman, 2003; Mansour et al., 1999) and these triple KS genes may be involved. *Hematodinium*, unlike the alveolate *Chromera velia* and diatom *Triceratium dubium,* has adenylation domains and condensation domains similar to the TeCATe transcript. Thus, it is possible that the triple KS and TeCATe transcripts were present in some form in the dinoflagellate common ancestor. They could either have been modified or lost such as in the *Amoebophyra* species that infect dinoflagellates and parasitize essential fatty acids from their host or kept intact in species like *Hematodinium* that infects crustacean hosts not known to make the polyunsaturated fatty acids found in dinoflagellates.

The origin of the single domain transcripts is much harder to ascertain due simply to the sheer number of very similar sequences. Previous studies have focused on phylogenies of adenylation and ketosynthase domains that could be annotated using traditional nomenclature (Beedessee et al., 2019; John et al., 2008; Kohli et al., 2016). This makes sense considering that these two domains represent the workhorse enzymes of modular synthesis and there is precedent for the gain or loss of a domain being diagnostic for toxicity (Kohli et al., 2015). Traditional nomenclature of polyketide synthases, however, is largely based on whether a gene is eukaryotic or prokaryotic and whether it is multi-domain or an assemblage of single domains, *i.e.*

type I and type II (Khosla, 2009), and dinoflagellates have been shown to possess genes similar to both eukaryotic and prokaryotic models that are both single and multi-domain (Van Dolah et al., 2017). Therefore, this nomenclature combined with distantly related model organisms such as humans and yeast is not very useful when trying to unravel the mechanisms underlying dinoflagellate modular synthases. Traditional nomenclature can also be misleading when trying to annotate sequences since protists are notoriously under-sampled in public databases relative to yeast, vertebrate models, and prokaryotes. In Kohli *et al.* 2017 (Kohli et al., 2017), 264 ketosynthase and ketoreductase single-domain transcripts as well as 24 multi-domain PKS transcripts were found in *G. excentricus* and *G. polynesiensis* transcriptomes using the BLAST2GO pipeline and HMMs based on annotations from previous studies. The present study using HMMs created from BLAST searches in A. *carterae* of the aforementioned domains in known multi-domain transcripts yielded 156 additional ketosynthase and ketoreductase domains in single and multi-domain transcripts for the same two *Gambierdiscus* species. More is not necessarily better, especially when confirming predictions experimentally is still out of reach. All possible genes could play a role in toxin synthesis ignoring bias from annotations in model organisms since atypical chemistry for model organisms appears to be the norm for toxin synthesis in dinoflagellates (Van Wagoner et al., 2014). This is also true of the BurA and ZmaK genes themselves that are atypical for prokaryote polyketide synthesis modules but appear to have been successfully transferred and retained in dinoflagellates.

The clustering analysis is in some ways more informative than phylogeny and annotation in that it gives an indication of the level of gene duplication for a domain within dinoflagellates and visualizes a large number of sequences without the preconception of a bifurcating evolution during speciation. It is also important to include as many domain types and not focus on ketosynthases alone since the loss of an acyl transferase or thioesterase can result in a truncated structure as hypothesized based on chemical comparison of pinnatoxin and gymnodimine to spirolides (Van Wagoner et al., 2014). The underlying hypothesis is that domains with sequence similarity may perform similar functions or be involved in similar pathways since this is often the primary constraint on evolution but neofunctionalization is also a possibility. Thus, the clusters were colored according to the presence of domains from the known multi-domain transcripts as a way of binning the clusters and begin to ascertain the functions of the many single domain transcripts. One example is the large number of these single adenylation and ketosynthase domain transcripts that are similar to the annotated transcripts as well as some of unknown similarity (Figure 1-11). Two reasonable explanations are that the domains themselves were serially duplicated and fragmented from parent multidomain transcripts resulting in gene expansion, or that there was a functional constraint forcing domains acquired by other means as single domain transcripts to evolve convergently and form multidomain transcripts. Possibly both are happening*, e.g.* gene duplication for the discrete cluster of BurA-like adenylation domains and convergent evolution for the ZmaK-like adenylation domains that are linked to another cluster of adenylation domains of unknown origin. For the thiolation domains convergent evolution is more likely considering that a single cluster encompasses domains from several different multi-

domain transcripts while the acyl carrier proteins have their own cluster (Figure 1-9). This would make sense considering that dinoflagellates only have between one and three phosphopantetheinyl transferases that can activate these thiolation domains (Williams, Bachvaroff, & Place, 2020). It is unclear if the ketosynthase domains from the central cluster containing multi-domain transcripts are performing similar functions and the outlying clusters are performing different functions such as chain length factors, or if ketosynthase domains are being acquired faster than convergent evolution is acting.

The acyl transferase and ketoreductase domain clusterings are especially interesting as the only case where the gene for fat synthesis is present in very large clusters with small clusters containing domains from the multi-domain transcripts (Figure 1-12). This was first described in the Symbiodiniaceae and described as FabD-like Trans ATs (Beedessee et al., 2019). While horizontal transfer and duplication of entire fat synthesis gene clusters has been shown (Chan, Baglivi, Jenkins, & Bhattacharya, 2013; Hutcheon et al., 2010), the extensive gene duplication suggested by the data presented here for the acyl transferase and ketoreductase genes would be unprecedented. Convergent evolution is unlikely given that fat synthesis is usually tightly regulated and none of the other fat synthesis genes show this type of clustering. It is possible that FabD and FabG like genes were coopted for some other function following an initial duplication. This would also indicate that these domains are performing a function separate from the multi-domain transcripts given that that they almost always have intact acyl transferase or ketoreductase domains. The triple KS is a special case here since the second ketosynthase is annotated as an acyl transferase containing ketosynthase by the conserved domain database of NCBI and the acyl transferase HMM only detected a domain in some transcripts but not others. This means that a Trans-acting acyl transferase is possible for some of the triple KS modules if the ketosynthase has lost the acyl transferase functionality, but this is speculation given these data. In general, the acyl transferase and ketoreductase clusters of unknown similarity were probably acquired later or gene duplication occurred in early dinoflagellates since they are in very low abundance in the basal species, *e.g. A. carterae* only has acyl transferases that are BurA-like (8 copies) and FabD-like (10 copies) and only 2 of 31 ketoreductases are found in the unknown cluster (Table S1-2).

The thioesterase and dehydratase clusterings paint a very different picture than the other abundant and diverse domains as one of the few cases where the domain count is consistently low with small clusters (Figure 1-10). There is still some gene expansion such as the Bur-A like thioesterases in *K. brevis* and *G. spinifera* that appear to have been duplicated along with other BurA-like domains, just to a lesser degree (Table S1-2). This small number of thioesterases in most species indicates that for very large toxins the number of synthetic complexes is low or that synthesis is highly iterative since a thioesterase is usually necessary to terminate each portion of synthesis (Khosla, 2009). However, it is important to remember that the "low abundance" of thioesterases is a relative description since thioesterases are more than nine fold more abundant in the core dinoflagellates than in the outgroup species (Table 1-2). The dehydratases, like the thioesterases, are usually encountered in multi-domain transcripts (Figure 1-3). Ketosynthase and ketoreductase domains on

the other hand are abundant as single domain transcripts. When looking at the chemical structure of many dinoflagellate toxins the acetate units are frequently hydroxylated, indicating that the ketone has been modified by a ketoreductase but not a dehydratase (Van Wagoner et al., 2014). These hydroxyls then frequently form epoxide bonds resulting in the "zipped up" structures of brevetoxin and yessotoxin. This makes the abundance of enoyl reductases strange since they would theoretically act after the dehydratases to further saturate the polyketide but the enoyl reductases are much more abundant than the dehydratases (Figure 1-10). It may be that many of the enoyl reductases have been coopted to operate on a substrate other than polyketides or that the dehydratases act as a chokepoint in synthesis and that their abundance is under tighter regulation or selection pressure. Given the number of enoyl reductase fragments (Figure 1-10B) it may also be that this gene is subject to a much higher level of gene duplication but that not all of the transcripts are being translated. Either way, the large number of enoyl reductases relative to dehydratases in dinoflagellates is in stark contrast to what is frequently described in prokaryote and fungal models where regulation of gene expression is much better understood and domain abundance directly correlates to the structure of the final product.

The phylogenetic analyses attempted on the thioesterase and acyl transferase domains had no bootstrap for all major nodes in spite of being able to produce clades with similar structure to the clustering output in the highest scoring trees (Supplementary files 1-1, 1-2). The only nodes with bootstrap support above 70% were those containing sequence variants from a single species or assembly variants from a single transcriptome. This is not surprising since gene copy number has made sequence phylogeny difficult for dinoflagellates in the past (Bachvaroff & Place, 2008; Bachvaroff et al., 2014; Janouškovec et al., 2017). Also, given the amount of horizontal gene transfer the concept of orthology become difficult to prove in general (Keeling, 2010), and in this case a functional approach is more useful if the goal is to extend hypothesis to biochemical characterization.

In general, there was a lack of condensation domains despite a large number of adenylation domains in all the core dinoflagellates. Although the condensation domains in the TeCATe transcript used to construct the HMM are similar to canonical condensation domains it is quite possible that there are other condensation domains not associated with multi-domain transcripts. It is certainly true that condensation domains can have their own specificity in natural product synthesis forming both amide and epoxide bonds without the aid of adenylation domains (Lin, Van Lanen, & Shen, 2009). Condensation domains are unlikely to play a large role in toxin synthesis in dinoflagellates given their almost ubiquitous use of acetate and general lack of amino acids although the frequent use of glycolate as a starter is conspicuous(Van Wagoner et al., 2014), and an unknown trans-acting condensation domain may be critical in initiating toxin synthesis.

## Scaffolding Domains and Single Domain Transcripts are Associated with Toxin Synthesis

Given their abundance, one could speculate that it is largely the single domain genes that are responsible for toxin synthesis with multi-domain genes like the triple KS responsible for the synthesis of poly-unsaturated fatty acids or portions of toxins like the acyl chains. I possibly these multi-domain genes or modules within them act on specific segments of toxin synthesis either individually or iteratively as has been proposed several times (Beedessee et al., 2019; Kohli et al., 2015; Van Dolah et al., 2017; Van Dolah et al., 2020). If it is mostly single domain genes involved in toxin synthesis then the thiolation domains with tetratricopeptide repeats may be important in scaffolding protein domains and providing reaction centers for the large complexes necessary to synthesize toxins (Clairfeuille, Norwood, Qi, Teasdale, & Collins, 2015). The fact that the fusion of a thiolation domain and a tetratricopeptide repeat is never found in conjunction with another domain and only present in the core dinoflagellates also correlates to toxin synthesis via single domain genes since none of the syndiniales or other alveolates transcribe large polyketides except *Hematodinium*, which possesses triple KS transcripts. Also, the acyl transferase domains may be involved in their own scaffolding or reaction center bridging given the occurrence of ankyrin repeats in many of the FabD like acyl transferase containing transcripts (Mosavi et al., 2004). The interplay between scaffolding by thiolation domains and reaction center bridging by trans-acting acyl transferases may be a driving force in the evolution of modular synthesis in dinoflagellates. Specifically, the number of acyltransferases relative to thiolation domains increases as one moves from the most basal Gymnodiniales to the more distal Gonyaulacales (Figure 1-5). This shift is also evident in the decrease in the mean number of thiolation domains in a transcript (Figure 1-6). The occurrence of multiple thiolation domains in tandem within a transcript was first observed in the *K. brevis* transcriptome and appears to be a hallmark of the Gymnodiniales that include species that make sterolysins and brevetoxin (Houdai et al., 2001; Ishida et al., 1995; Meng et al., 2010; Peng, Place, Yoshida, Anklin, & Hamann, 2010; Van Wagoner et al., 2008)as well as the Suessiales that can make zooxanthellatoxin and zooxanthellamide(Fukatsu et al., 2007). Unfortunately, this is not diagnostic with many species that do not have a described toxin such as *Pelagodinium beii* having a higher average thiolation domain (1.47) than *Karlodinium veneficum* (1.28) that makes karlotoxin. The *Gambierdiscus* species had the highest average number of thiolation domains among the Gonyaulacales (1.58 and 1.68) that otherwise had low averages indicating that multiple thiolation domains may be one strategy in synthesizing long polyketides such as ciguatoxin (Satake et al., 1997). Unfortunately, many polyketides in dinoflagellates are likely undescribed if they do not impact humans making it difficult to correlate the synthesis of dinoflagellate polyketides to molecular results.

## Conclusion

In general, there was no overarching signal relating domain count or domain expansion to toxin production as shown in the principal components plots (Figure 1-4). This is largely due to the abundance of modular synthase genes among all core dinoflagellates investigated. So, if core dinoflagellates are all making polyketides, what is their purpose? *Karlodinium veneficum* is the only case where an ecological role has been identified, *i.e.,* prey capture (Sheng et al., 2010), but it is also the only case where the toxin is found readily outside the cell. Another role that has been identified is mediating redox potential in the chloroplast of *Karenia brevis* (Chen et al., 2018). This helps explain why complex polyketides are found in photosynthetic species as well as the apparent association between function and synthesis in the chloroplast since it is a major source of redox stress. Thus, focusing on "toxin" synthesis may not be advantageous in the long run versus understanding the modular synthases in dinoflagellates as a whole and their biological role within dinoflagellates. Just as subtle differences in the availability of a thioesterase or acyl transferase can radically alter the final structure of a polyketide, assays that identify known toxins can falsely label a species or strain as being non-toxic despite that organism making polyketides that are only subtly different than the toxin standards.

The data presented here shows long-term evolution along the entire scope of dinoflagellate history with the acquisition of tetratricopeptide repeats fused to thiolation domains in the core dinoflagellates and the increase in acyl transferase domains as a major component of the synthetic domain population, specifically the FabD-like acyl transferases. Short-term evolution with rapid increases in the copy number of certain domains that was first shown in *Symbiodinium* species (Beedessee et al., 2019) appears to be a universal feature of dinoflagellate evolution that could also explain why many of the larger toxins are unique to certain lineages. While it seems like a natural progression to use molecular datasets from dinoflagellates to make predictions about the functionality of synthetic domains, existing datasets have been validated with species very distantly related to dinoflagellates, and protists in general, making these predictions unlikely to be realistic. For example, the Beedessee et al. paper from 2019(Beedessee et al., 2019) used up to date methods to predict the substrates of adenylation domains in dinoflagellates resulting in tryptophan, phenylalanine, and glycine. This is unlikely to be true since tryptophan and phenylalanine are not found in described dinoflagellate natural products that would utilize adenylation domains. Also, using the same method as the Beedessee et al. paper to predict the substrate for the *A. carterae* adenylation domain of BurA, as well as from the original BurA sequence from *Burkholderia*, similarly results in phenylalanine, but this was shown to actually be a methionine modified to a propanal by radioisotopic labeling in the bacterium (Franke et al., 2012)

# Chapter 2: The Phosphopantetheinyl Transferases of Dinoflagellates

*Abstract*

Dinoflagellates play important roles in the world's ecosystem in carbon capture and recycling as well as the production of polyunsaturated fatty acids. They make natural products that can harm environmental and human health, but these products are also possible therapeutics with chemistries that cannot currently be synthesized in the lab. These natural products as well as lipids are synthesized in a modular fashion on one or a series of carrier domains with each piece of the product added to the carrier and modified in series. Results and interpretation from the previous chapter demonstrated massive numbers of putative toxin synthesis domains in individual dinoflagellate transcriptomes with no apparent correlation to toxin production. However, it was evident that the thiolation domain that acts as a carrier to scaffold synthesis could be binned into two distinct groups likely representing lipid and natural product synthesis. The first rate-limiting step in synthesis requires the addition of the phosphopantetheinate arm from CoA to the carrier that provides a free thiol upon which the synthetic units are added and removed. This attachment of the phosphopantetheine is performed by a phosphopantetheinyl transferase (PPTase) that is usually specific for either an acyl carrier protein used in lipid synthesis or a thiolation domain used in the synthesis of other natural products. In this study the PPTases of dinoflagellates were enumerated, their expression patterns characterized, and their sequence analyzed for motifs to help explain their biological roles using the basal toxic dinoflagellate *Amphidinium carterae* as a model. The acyl carrier protein expression was also characterized during a growth curve with the ultimate goal of identifying PPTases used in lipid synthesis versus natural product synthesis. Two of the three PPTases showed an alternating expression in a day night cycle as well as during a growth curve with variable expression depending on the growth stage. The final PPTase was never observed in its whole form and this PPTase was found to not have a stop codon. Based on western blots, all three had a noticeable breakdown product that appears to cleave between two helices rendering the protein non-functional. None of the PPTases had an expression pattern or a chloroplast targeting sequence like what was observed for the acyl carrier protein indicating that the role of the PPTases in dinoflagellates may be multi-functional.

*Introduction*

The 4′-Phosphopantetheinyl transferases (PPTase) are responsible for the post-translational modification of carrier proteins in many primary and secondary metabolic pathways (Beld et al., 2014; Lambalot et al., 1996). The carrier proteins can be stand-alone proteins such as the bacterial acyl carrier proteins (ACP) in fatty acid synthesis, or domains within multifunctional proteins such as the ACP or peptidyl carrier protein (PCP) domains in polyketide synthases (PKS) and non-ribosomal peptide synthetases (NRPS), respectively (Bentley & Bennett, 1999; Khosla, 2009). PPTases transfer the phosphopantetheinyl group of co-substrate CoA to a conserved serine residue in carrier proteins creating a free thiol group. The modification of carrier proteins with the flexible phosphopantetheinyl group allows them to shuttle acyl intermediates between domains through reversible formation of a thioester linkage. This allows for the normally modular synthesis (Khosla et al., 2009) of many natural products including antibiotics and polyunsaturated fatty acids (Gurney & Thomas, 2011; Yazawa, 1996) while saturated lipids are synthesized with the same biochemistry but in an iterative fashion with a single carrier protein (Buhman et al., 2001; Hölzl & Dörmann, 2019).

Dinoflagellates are marine protists that can be readily split into two evolutionary groups, the basal syndiniales that are heterotrophic parasites and the distal "core" dinoflagellates (Bachvaroff et al., 2014; Hoppenrath & Leander, 2010; Janouškovec et al., 2017) that are frequently mixotrophic with a complex evolutionary history with multiple chloroplast acquisitions and losses (Cavalier-Smith, 2002; Dorrell & Howe, 2015; Ishida & Green, 2002; Janouskovec et al., 2010; Yamada et al., 2017; Yamada et al., 2019). The photosynthetic species, like other algae, produce many light sensing compounds, but they are also a source of many other natural products such as the polyunsaturated fatty acid docosahexaenoic acid (DHA) and compounds that can block ion channels associated with nerve function or create pores in membranes containing cholesterol (Javed et al., 2011; Mansour et al., 1999; Wang, 2008). Just as in bacteria and fungi, the majority of dinoflagellate natural products are made in a modular fashion whereby a carboxylic acid, usually acetic acid but sometimes other small carboxylic acids or an amino acid, is attached to a carrier protein and chemically modified followed by another addition and modification and so on (Bentley & Bennett, 1999; Izoré & Cryle, 2018; Jenke-Kodama & Dittmann, 2009; Khosla et al., 2009; Khosla, 2009; Wang et al., 2014). This is a deceptively simple means of biosynthesis, similar to protein synthesis, but with an almost limitless number of substrate and modification combinations giving rise to a huge diversity of compounds (Javed et al., 2011). Even when limiting the search to dinoflagellates, many compounds have been discovered (Van Wagoner et al., 2014), usually noticed due to their impacts on human and animal health (Twiner et al., 2012; Walsh et al., 2015; Wang, 2008), leaving room for many more to be discovered. As with all natural product synthesis the attachment of a phosphopantetheinate group to the carrier protein is required to provide a labile substrate. The PPTase that performs this operation is considered vital for life since it acts in lipid synthesis (Beld et al., 2014). This is not true for many alveolates that parasitize host organisms and have a limited capacity for lipid synthesis (Leblond &

Dahmen, 2012; Mazumdar & Striepen, 2007). In general, PPTases are specific for each pathway that they activate with some examples of generalism in natural product synthesis (Gerc, Stanley-Wall, & Coulthurst, 2014) but always a separation of lipid and natural product synthesis with the green alga *Chlamydomonas reinhardtii* being a rare exception (Sonnenschein, Pu, Beld, & Burkart, 2016).

The goal of this study was to characterize the PPTases of dinoflagellates to determine if they could be functionally binned into lipid and natural product synthesis based on their sequence and expression patterns. *Amphidinium carterae*, a basal toxic dinoflagellate was used as a model for sequence retrieval and for monitoring protein expression. The three PPTase sequences from *A. carterae* were each placed in a distinct dinoflagellate clade. Their protein expression was compared to the acyl carrier protein, the lipid synthesis thiolation domain, to determine if one of the three had a correlative expression pattern indicating a role in lipid synthesis. The results were quite atypical with an alternating expression of two of the PPTases, both over 12 hours as well as during growth, while the final PPTase was never expressed in its full form at all. This implies that the PPTases may not have a dedicated function and that some copies may be redundant. This helps to explain why members of some clades are frequently lost during dinoflagellate evolution.

## *Materials and Methods*

### Sequence collection, analysis, and construct generation

Transcriptomes of dinoflagellates species used in this study were assembled using Trinity v2.3.2 from sequences deposited in the Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis (CAMERA) (Sun et al., 2011) database as described in Janouškovec *et al.*, 2017. *Amphidinium carterae* (Hulbert 1957) was used as the representative species among "core" dinoflagellates because it is the most basal toxin producing species and has a small genome. Candidate phosphopantetheinyl transferases (PPTases) and acyl carrier protein were retrieved from the *A. carterae* transcriptome using annotations from the BLASTX (Altschul, Gish, Miller, Myers, & Lipman, 1990) results against the non-redundant protein database at the National Center for Biotechnology Information (NCBI). The three PPTase sequences were then modeled against available crystal structures for PPTases using the protein homology/analogy recognition engine (Phyre v2.0) to confirm the annotation based on conserved structure and residues from Beld et al. 2014. PPTases from other dinoflagellates species were retrieved using reciprocal TBLASTX against the assembled transcriptomes with each of the *A. carterae* sequences as query. A reciprocal approach was used to minimize spurious sequences whereby the subject sequence of each search was used as a query and was kept only if it produced the original query sequence as the top BLAST hit in its own search. Outgroup sequences were obtained from NCBI's Genbank from *Homo sapiens* and several other fungi and bacteria known to produce secondary metabolites via

polyketide synthases. All sequences from selected species that were annotated as phosphopantetheinyl transferases were used as candidate sequences in initial alignments. Trees were constructed from the resultant alignment using RAxML v8.2.12 (Stamatakis, 2014) with a gamma distribution and invariant site estimations using the WAG substitution model. Rapid bootstrapping was employed with 100 replicates and random seed 11111. Long branches from the resultant tree were assessed in the alignment and removed if it was determined that the sequence was truncated or likely to be a spurious annotation based on conserved residues. This was done iteratively until an increase in bootstrap scores was no longer attainable.

To obtain full length sequences of the *A. carterae* PPTases, primers were designed to amplify the 5′ and 3′ untranslated regions (UTRs) based on the open reading frames of each PPTase along with the spliced leader (Lidie & Van Dolah, 2007; Zhang et al., 2007), a low variability sequence spliced onto each messenger RNA in dinoflagellates, and a poly-T primer with a GC lock and a priming sequence not found in the *A. carterae* transcriptome (Table 2-1). RNA was

| Primer Name | Sequence 5′ to 3′ | Length | Annealing Temp. °C |
|---|---|---|---|
| PPT1_CDSF4 § | GCTTACAGTGGAGGCCCTACTTCCAATGGG | 30 | 74.9 |
| PPT2_CDSF4 § | TCCCTGCGGTGTCCAACTTCAAGCTTTACA | 30 | 72.1 |
| DTR2 § | CATCTTGCTAGCTCGCGATCTTGAAGTAGTC | 31 | 72.1 |
| Dino_SL § | TCCGTAGCCATTTTGGCTCAA | 21 | 59.5 |
| anchoredDTR2 ♀ | CATCTTGCTAGCTCGCGATCTTGAAGTAGTCTTTTTTTTTTTTTTTTTTS | 52 | 41.9* |
| Primers listed with an "§" were used in PCR amplification while those with an "♀" were use in reverse transcription | | | |
| The "anchored DTR2" annealing temperature denoted with a "*" is for the poly-T region only. | | | |

isolated from *A. carterae* cultures using Tri-reagent (Sigma Aldrich, St. Louis, MO) according to the manufacturer's directions and reverse transcribed using Superscript II reverse transcriptase (Thermo Fisher) with 50 nM poly-T primer and and 5 nM of each reverse primer from the PPTase open reading frames according to the directions of the reverse transcriptase. Amplification of cDNA template was performed using the Phusion high fidelity polymerase (New England Biolabs, Ipswich, MA) with final concentrations of 1 ng/µl of template and 500 nM of forward and reverse primers for each PPTase for each reaction. Thermal cycling conditions consisted of an initial denaturation of 98 ℃ for 2 minutes followed by 35 cycles of denaturation at 95 ℃ for 15 seconds, annealing at 60 ℃ for 5′ UTR reactions and 68 ℃ for 3′ UTR reactions for 20 seconds, and extension at 72 ℃ for 1 minute and a final polishing step at 72 ℃

for 5 minutes. This resulted in several bands when visualized on a 1% agarose gel in 0.5× TBE separated at 15 V/cm. Gel excision was performed using the Monarch Gel Extraction kit from New England Biolabs and the bands were sequenced on an Applied Biosystems 3130XL fragment analyzer at the Bioanalytical Services Lab (BasLab) in Baltimore, MD. The 5′ UTR for Clade 2 and three PPTases were sequenced confirming the spliced leader as well as the 3′ UTR of the Clade 3 PPTase. The 5′ sequence of the Clade 1 and 2 PPTases were very similar and indicated that the Clade 1 PPTase sequence started just after the spliced leader negating a need for further sequencing. The 3′ ends of the Clade 1 and 2 PPTases were very difficult to sequence and a BLAST analysis of the primers used against the *A. carterae* transcriptome showed many sequences with very high or identical similarity, despite a length of twenty bases. A thirty base length primer set was designed (Table 2-1) and the amplification and sequencing methods were attempted again resulting in 3′ sequence for both the Clade 1 and 2 PPTases. Full length sequences were deposited in Genbank (Accession #ON157050-ON157052 for PPTase Clade 1, 2, and 3, respectively) and used for further analyses. Predicted folding structure of each of the PPTase 3′ UTRs was performed at the Fold Web Server (https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Fold/Fold.html visited 12/15/2021). Subcellular localization motifs were predicted using WolF PSORT (https://wolfpsort.hgc.jp/ visited 02/09/2021) with the animal sequence database as well as SignalP (https://services.healthtech.dtu.dk/service.php?SignalP-4.1 visited 02/09/2021). Ubiquitination site prediction was performed with UbiSite (http://csb.cse.yzu.edu.tw/UbiSite/ visited 01/25/2021) using the high threshold cutoff. Protein stability, molecular weight and PI were calculated using the ProtParam program (https://web.expasy.org/protparam/)

The open reading frames of the three *A. carterae* PPTases (a stop codon was assumed for PPTase 2) and the acyl carrier protein were used to generate protein expression constructs using the commercial services from Genscript (Piscataway, NJ, USA). Amino acid sequences were optimized for expression in *E. coli* and placed into a pET-20b vector (Supplementary File S2-1). Epitopes were also determined and antibodies produced in rabbits (Table 2-2) for western blotting.

| Gene | Epitope | Molecular Weight (kD) | Isoelectric Point |
|---|---|---|---|
| PPTase Clade 1 | CAAPQLERGEGEDLS | 39.5 | 5.24 |
| PPTase Clade 2 | CVRQEGSLPARYEGA | 39.5 | 7.98 |
| PPTase Clade 3 | KGDRLHYKLSKGSGC | 44.1 | 6.82 |
| ACP | EEFEVDLPDEETTELKN | 13.2 | 4.09 |
| All sequences are from *Amphidinium carterae* | | | |

*Amphidinium carterae* growth curve and gene expression

For diel expression of PPTases *Amphidinium carterae* strain NCMA 1314 was grown axenically in L1 EH1(Berges, Franklin, & Harrison, 2004) medium without silicate modified to have 1 mM HEPES and with 100 µg/ml carbenicillin, 50 µg/ml kanamycin, and 50 µg/ml spectinomycin at 20° C and 14 hours of light at approximately 50 µmol of photons cm-2 s-1. The culture was split into thirteen duplicate 25 cm2 vented flasks and a whole flask was taken at each timepoint over a 12 hour period. Starting at 6 hours before lights out a sample was taken every two hours until 2 hours before lights out, after which a sample was taken every 30 minutes until two hours after lights out returning to a sample every two hours. Each sample was split into two 50 ml tubes and centrifuged at 1000 x g for 10 minutes to collect the cells. This entire process was repeated with a new culture and samples were takemn every 2 hours over a 24 hour period. One half of each sample was suspended in 2x SDS PAGE loading buffer (80mM Tris pH 6.7, 2% sodium- dodecyl sulfate, 10% glycerol, 1mM dithiothreitol, and 6 ppm bromophenol blue) while the other half was suspended in Tri-Reagent (Sigma T9424). RNA was extracted from the Tri-Reagent fraction of the 12 hour sampling according to the manufacturer and cDNA was generated from 1 µg of total RNA using random primers (Invitrogen 48190011) and Superscript II (Invitrogen 18064-022) according to the manufacturer's directions. Relative quantification of transcripts was determined using the primers in Table 2-3

| PPTase | qPCR Forward | qPCR Reverse |
|---|---|---|
| Clade 1 | TTGCCAGAAGCAGACAGAGA | AAGTTGGGCATACGATCTGG |
| Clade 2 | GTGATTGGGTCGTTCTTGCT | TGGAAGGCCTCATAGAGCAT |
| Clade 3 | TCGGCATTGATGTAGCAGAG | CATCCCCTCTAGCTTTCACG |

and primers for rpl7 (Jones et al., 2015) as a normalizing control at 500 nmol and the iTaq Supermix with ROX (Bio-Rad 1725121) according to the manufacturer's directions with the following thermal cycling parameters in an Applied Biosystems 7500 Fast Real Time PCR machine: Initial denaturation at 95° C for 2 minutes followed by forty cycles of denaturation at 95° C for 15 seconds and annealing and extension at 60° C for 30 seconds. Data was acquired during the annealing and extension stage and the reaction was followed by a melt curve to test for spurious products. The resultant cycle thresholds (Cts) were subtracted from a common zero template value of 35 cycles to give inverse Cts for visualization.

Protein abundance of PPTases during diel expression was determined by western blotting using antibodies generated by Genscript made from the epitopes listed in Table 2-2. For each timepoint, 15 µl of samples in 2x SDS PAGE loading buffer were loaded onto a 4-12% Bis-Tris gel (Novex NP-0323) and run in MOPS buffer at a constant voltage of 165 for 50 minutes. Peptides used to generate the antibodies for each other PPTase were also loaded as a negative control. The gel was transferred onto a PVDF membrane using a Bio-Rad Trans-Blot Turbo using the high molecular weight transfer protocol according to the manufacturer's directions. The blot was probed using the Novex iBind Western system with a 1:400 dilution of the

1° antibody and a 1:400 dilution of the goat-anti rabbit HRP conjugated 2° antibody according to the manufacturer's protocol. The probed blot was exposed to the Bio-Rad Clarity Western ECL substrates for 5 minutes and imaged using a Bio- Rad ChemiDoc Touch imaging system with optimal exposure. Relative quantification of bands was determined using Image Lab software (Bio- Rad v5.2.1).

For determining PPTase and acyl carrier protein expression during a growth curve a 50 ml culture of *A. carterae* was started in the same manner as with diel expression profiling. Upon reaching approximately 30,000 cells/ml the culture was diluted 1:2 every Monday, Wednesday and Friday until a volume of 1 L was reached. The culture was then transferred into a 20 L multiport polycarbonate vessel with aeration supplied through a 0.2 µm filter. Dilution continued every Monday, Wednesday and Friday using sterile media without antibiotics to a volume of 18 liters and a density of 30,000 cells/ml. pH was maintained at 7.8 ∓ 0.2 using a pH controller with a solenoid attached to a CO2 cylinder bubbling at a rate sufficient to correct the pH by 0.2 units in approximately 1 minute. 500-1000 ml of high density and low-density culture, respectively, was harvested by centrifugation at 1000 × g for 10 minutes at 4 ℃ on days 0, 2 5, 7, 9, 12, 14, 16, 19, 21, 23, and 26. Cell counts were taken and the cell pellet was diluted in 2× SDS-PAGE sample buffer to a concentration of 40,000 cells/µl. To generate protein standards for semi-quantitation The *E. coli* BL21(DE3) cells containing pET-20b plasmids with each of the three codon optimized PPTase sequences were grown in autoinduction media (Studier, 2005) with 100 µg/ml carbenicillin at 25 ℃ for three days at 250 rpm. Induced *E. coli* clones were collected after protein expression by centrifugation at 10,000 x g for 15 minutes at 4 ℃ and the supernatant was decanted. Pellets were stored at -80 ℃ until processing. Frozen pellets were suspended in 25 ml of lysis buffer containing 5 mM imidazole, 500 mM NaCl, 25 % (v/v) glycerol and 20 mM Tris/HCl pH 7.5 along with bacterial protease inhibitors (Sigma Aldrich, St. Louis MO) and thawed at 4 ℃. Cells were lysed with a French press chilled to 4 ℃ with 1000 PSI at the piston and 20,000 PSI at the outlet. The lysate was clarified by centrifugation at 12,000 x g for 15 minutes at 4 ℃ and the supernatant was treated with benzonase (Thermo Fisher) overnight at 4 ℃. The His-tagged lysate was bound to a 1 ml Hi-trap crude cobalt column (Cytiva); washed with 25 column volumes of 5 mM Imidazole, 250 mM NaCl, 20 mM Tris/HCl pH 7.5; and eluted into 10 separate volumes of 250 mM Imidazole, 250 mM NaCl, 20mM Tris/HCl pH 7.5, and 12.5% glycerol using an AKTA chromatography system (Cytiva) (Supplementary Figure S2-1). Elution fractions were separated with 4-12% Bis-Tris gels from Novex (Thermo Fisher) in MOPS buffer at a constant voltage of 165 V for 50 minutes and stained with Imperial stain (Thermo Fisher) according to the manufacturer's directions to verify protein capture. The purified proteins were also separated by SDS-PAGE electrophoresis, blotted, and probed in the same manner as diel expression starting with 100 nM estimated protein followed by seven 3-fold dilutions to verify PPTase production and establish a standard curve (Figure 2-1). This was not done for the

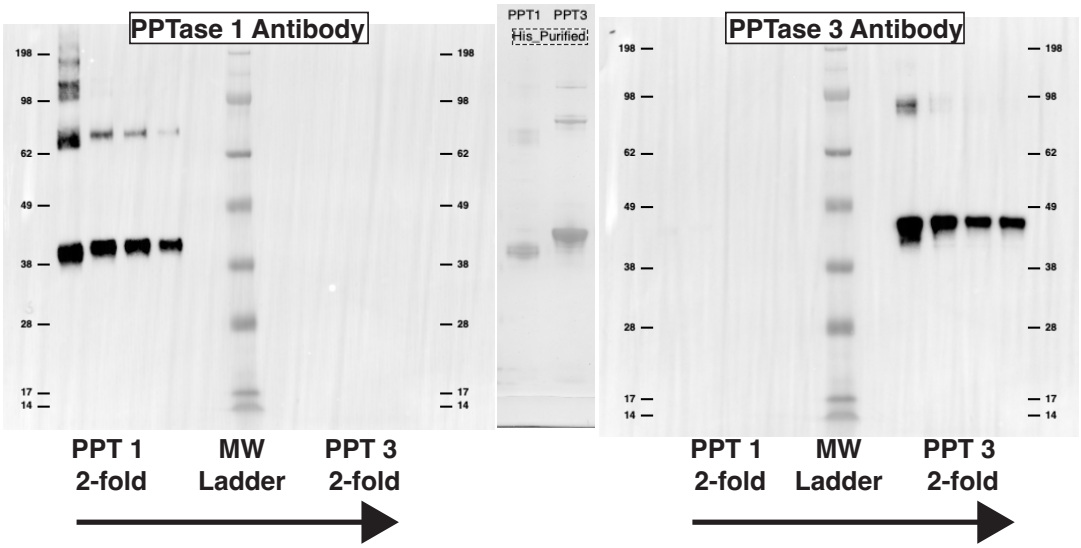



**Figure 2-1:** ***Purified protein controls for western blotting***
Purified protein of *Amphidinium carterae* phosphopantetheinyl transferases (PPTase) used as western blotting standards are shown. The His-tagged purified proteins are shown in the middle as a coomassie stained gel scaled to the same dimensions as the western blots. Each western blot has a 2-fold dilution of PPTase 1 on the left and PPTase 3 on the right with a molecular weight marker in the middle. The left blot was imaged with a PPTase 1 primary antibody while the right blot was imaged with a PPTase three primary antibody.

Clade 2 PPTase because soluble protein was not produced using this method.

A total of 100 ng purified protein for quantification purposes and 10 µl (400,000 cells) of total protein from each growth curve time point along with a Seeblue plus2 pre-stained ladder were separated on a 4-12% Bis-Tris gel (Novex, Waltham, MA) in MOPS buffer at 165 V for 50 minutes for western blotting with antibodies to each of the three PPTases and the ACP (Table 2-2). Separated proteins were transferred to PVDF membranes using the Trans-Blot Turbo Transfer system from BioRad (Hercules, CA) according to the manufacturer's settings for a standard 1 mm gel. Total protein transferred was quantified using the AzureRed total protein stain from VWR (Radnor Township, PA) and imaged on an Azure imaging system according to the manufacturer's protocols. Blots were blocked and exposed to primary and secondary antibodies using the iBind system (Thermo Fisher, Waltham, MA) with a 1:500 dilution of each primary antibody (Genscript, Piscataway, NJ) and a 1:50,000 dilution of a horseradish peroxidase conjugated goat anti-rabbit secondary antibody (BioRad). Westerns were imaged using the SuperSignal West Pico chemiluminescence kit from Thermo Fisher on a BioRad ChemiDoc system with optimal exposure and 4×4 binning. The pre-stained ladder was also imaged and the two images merged in BioRad's Image Lab software version 6.1.0. Bands were identified based on relative molecular weight and quantified based on band density. The conversion of band density to nanogram estimates was based on a comparison of the 100 ng standard from each blot to a standard curve using 3-fold dilutions of purified protein and a power equation describing the relationship between concentration and band density from *https://www.dcode.fr/function-equation-finder*

that was modified to better fit data at lower concentrations. This was not done for the Clade 2 PPTase since the full-size protein was never observed.

## *Results*

### Phosphopantetheinyl transferase phylogeny

A final alignment was made of phosphopantetheinyl transferases from 38 species of dinoflagellates among 45 transcriptomes including three transcriptomes from co-infections of a core dinoflagellate and a dinoflagellate parasite of the genus *Amoebophyra*, and the non- photosynthetic species *Oxhyrris marina*, *Noctiluca scintillans*, and *Crypthecodinium cohnii.* Dinoflagellate sequences coded for a predicted helical and sheet secondary structure described as a hallmark of PPTase amino acid sequences (Beld et al., 2014). Non-dinoflagellate sequences annotated as PPTases included *Chlamydomonas eustigma* (2 sequences), *Chlamydomonas reinhardtii* (1 sequence), *Homo sapiens* (1 sequence), *Phellinus noxius* (3 sequences), *Sterium hirsutum* (4 sequences), *Punctularia strigosozonata* (2 sequences), *Streptomyces venezuelae* (11 sequences), *Streptomyces lividans* (7 sequences), *Streptomyces laurentii* (1 sequence), and *Streptomyces lavendulae* (12 sequences). The total alignment length was 976 characters including gaps. The resultant tree (Figure 2-2) placed all dinoflagellate sequences (excluding *Oxhyrris*

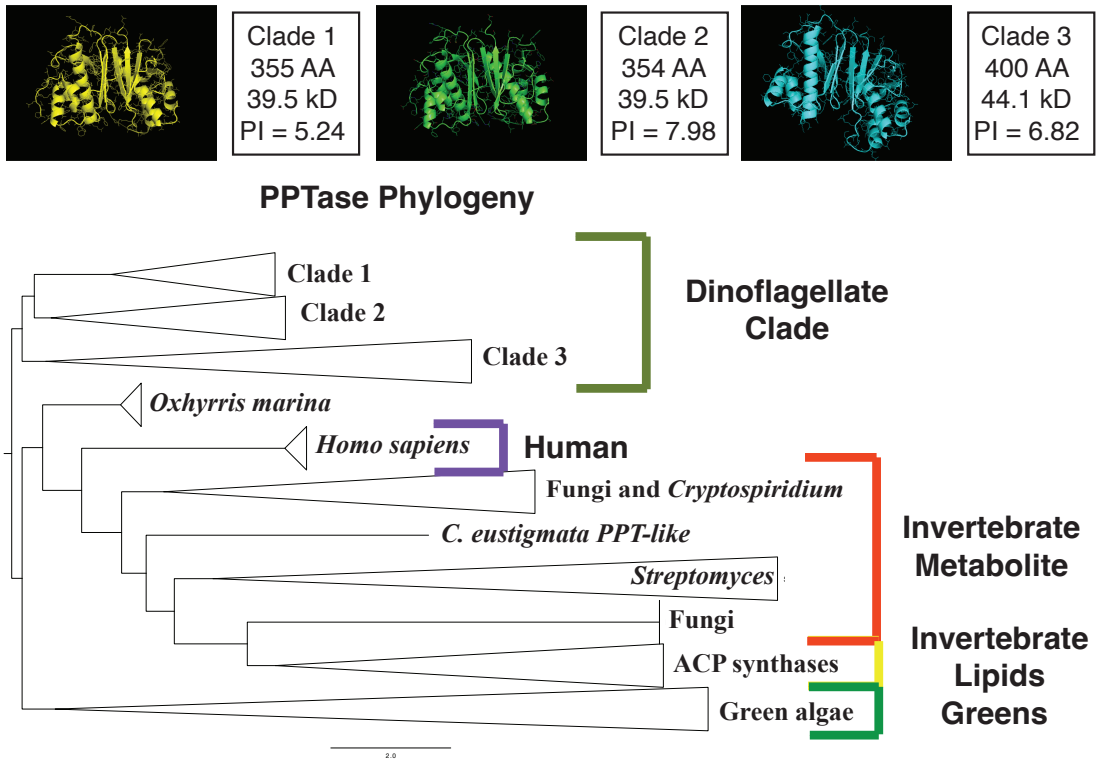



**Figure 2-2: Dinoflagellate phosphopantetheinyl transferase phylogeny.**
A collapsed phylogenetic tree is shown for phosphopantetheinyl transferases (PPTases) from dinoflagellates, green algae from the genus *Chlamydomonas*, human, and several bacteria and fungi. Clades have been collapsed with the size of the triangle equal to the total branch lengths. Groupings have been color coded for functional and/or taxonomic groupings and labeled. Shown above the tree are folding patterns and characteristics for the *Amphidinium carterae* PPTases from each of the dinoflagellate clades.

*marina* which formed its own clade) outside of all other clades with 68% bootstrap support (Supplementary File 2-2). Within the dinoflagellates there were three clades with poor bootstrap support. These were arbitrarily named Clade 1 (*Amphidinium carterae* sequence comp10839_c0_seq1, 24% bootstrap support), Clade 2 (*Amphidinium carterae* sequence comp29939_c0_seq1, 52% bootstrap support), and Clade 3 (*Amphidinium carterae* sequence comp25404_c0_seq1, 32% bootstrap support). Using each of the representative sequences from *A. carterae* as a query in a BLAST search of dinoflagellate transcriptomes resulted in the retrieval of most if not all of each species' PPTase sequences, indicating that the low booststrap support is due to high sequence similarity among the PPTase clades. Removal of the dinoflagellates lowered the bootstrap support for most outgroup clades except for the three *Chlamydomonas* sequences that approximately doubled their bootstrap support (tree not shown). Dinoflagellate Clade 3 PPTases contained all species examined while Clade 2 and 1 contained 30 and 27, respectively. All species contained at least two PPTase isoforms with the exception of *Protoceratium reticulatum* that only contained a Clade 3 sequence despite a robust transcriptome. *O. marina* also only appears to have one PPTase that was placed outside of all dinoflagellate clades. For *A. carterae PPTases*, sequence comparision found 85 conserved residues with

pairwise similarities of 39.1%, 37.7% and 45.4% for Clade 1 versus 2, Clade 1 versus 3 and Clade 2 versus 3, respectively.

Phosphopantetheinyl transferase expression patterns during growth

Over the twelve hour growth period the transcript abundance was flat for all three PPTases as well as the ribosomal protein except for a period immediately after lights off when the abundance of all transcripts dipped and then returned to previous values (Figure 2-3). The protein abundance was very different with the Clade 1

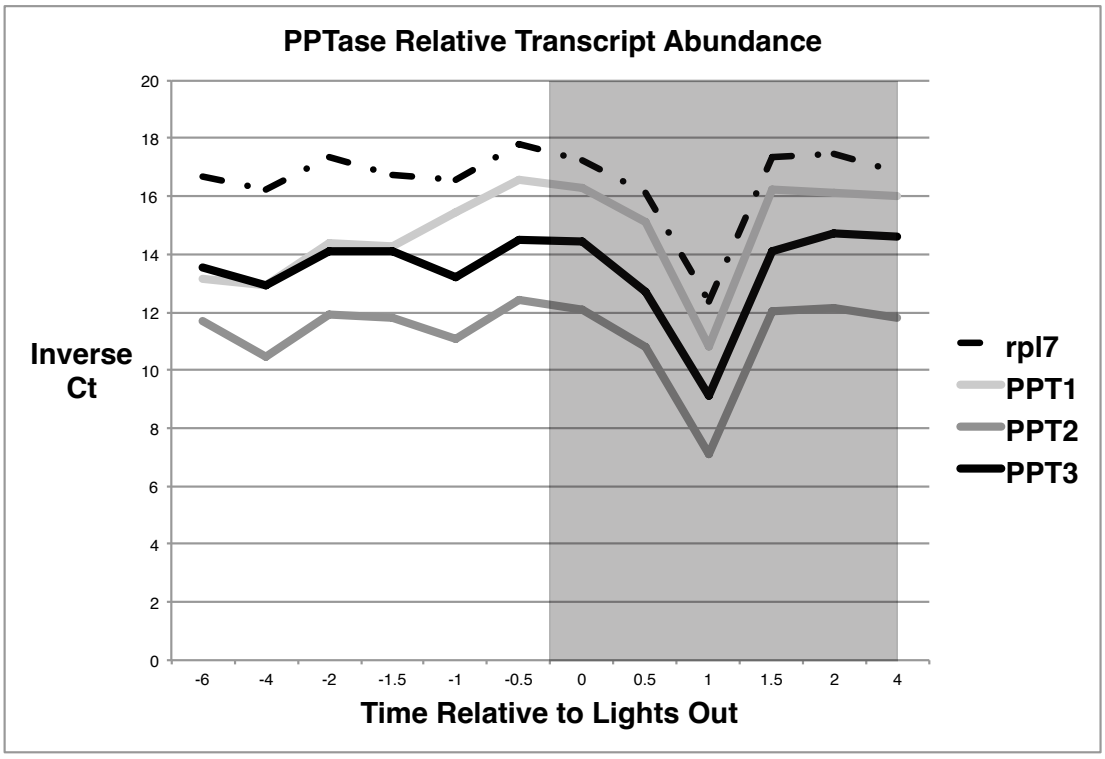



**Figure 2-3: *Amphidinium carterae* transcript abundances over a 12-hour period.** Transcript abundances based on qPCR for four *Amphidinium carterae* genes are shown over a 12 hour period as an inverse cycle threshold (Ct) on the Y-axis and time relative to incubator lights out on the X-axis. The genes are ribosomal protein l7 as a broken line, Clade 1 phosphopantetheinyl transferase (PPTase) as a light gray line, Clade 2 PPTase as a dark gray line, and Clade 3 PPTase as a black line. The graph is shaded on the right side to indicate lights off.

PPTase showing peak expression when the lights turn off and no expression at the first two time points whereas Clade 3 was only expressed at the first two time points. The Clade 2 PPTase is expressed at all time points with a similar pattern to PPTase 1 but only as a small band (Figure 2-4). All three PPTases showed this lower sized band at

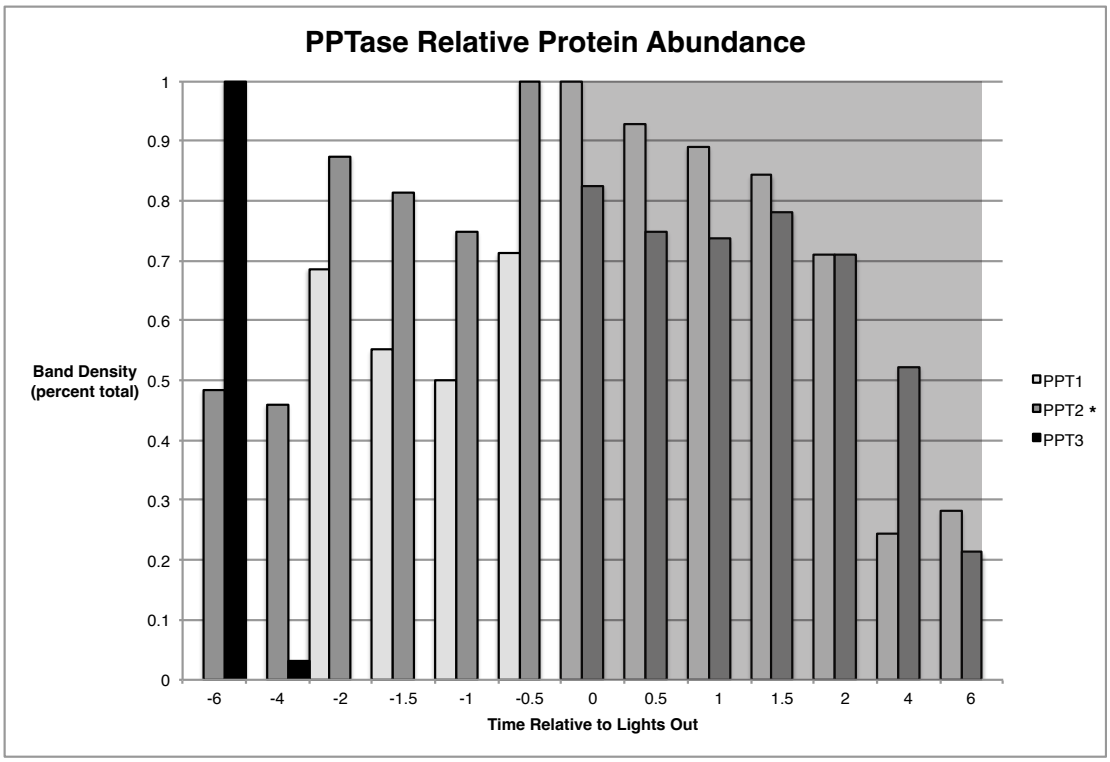



**Figure 2-4: *Amphidinium carterae* protein abundances over a 12-hour period.** Relative Protein abundances based on western blotting for three *Amphidinium carterae* phosphopantetheinyl transferases (PPTases) are shown over a 12 hour period as a percent of the darkest band on the Y-axis and time relative to incubator lights out on the X-axis. The Clade 1 and 3 PPTases are shown as a light gray and black bar, respectively, and represent the full sized band. Clade 2, highlighted with a "*" is shown as a dark gray bar but is the lower band common to all three PPTases.

various time points, including the Clade 3 PPTase following a loss of the full sized band in the first diel experiment (Figure 2-5). The expression patterns also differed

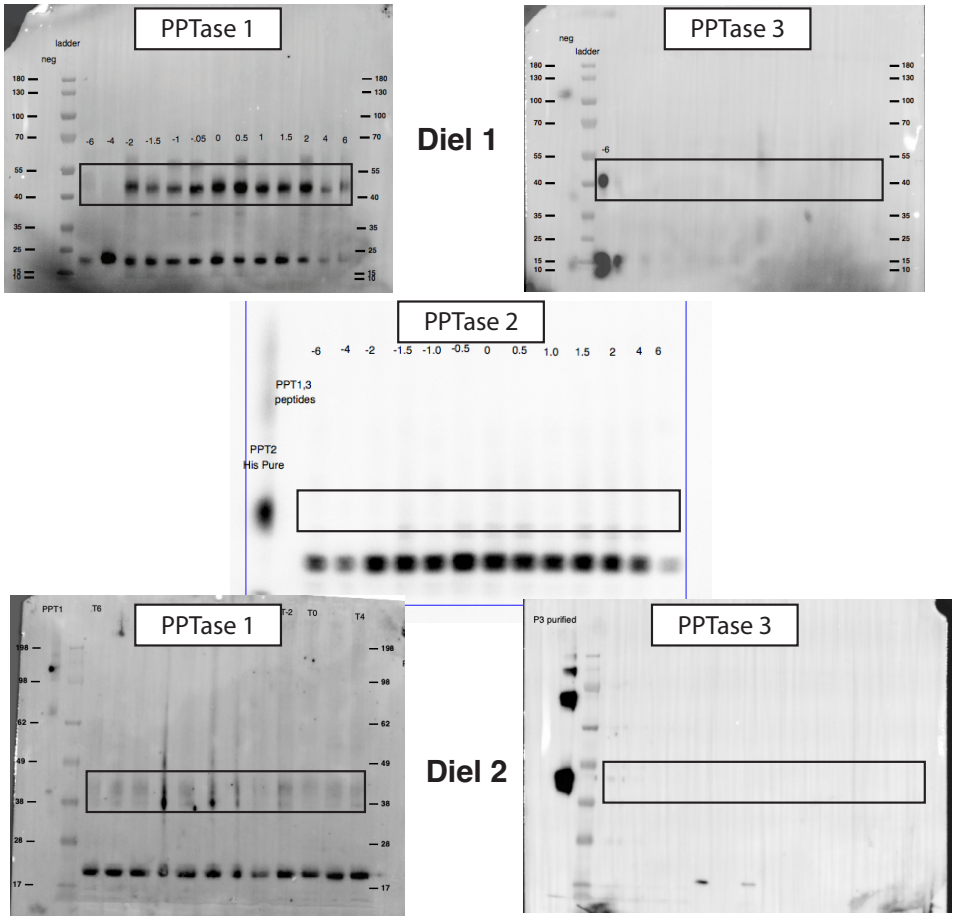



**Figure 2-5: *Amphidinium carterae* protein abundances over a 12-hour period in replicates cultures.**
Western blots for three *Amphidinium carterae* phosphopantetheinyl transferases (PPTases) are shown over a 12 hour period. The top two westerns are from the first diel experiment with the Clade 1 phosphopantetheinyl transferase (PPTase) antibody on the left and Clade 3 on the right. The middle panel shows the Clade 2 antibody for the first diel. The bottom two panels are from the second diel experiment also with Clade 1 and 3 PPTase antibodies. The expected size in marked by a black box in all images. For the images where the expected size is not apparent, recombinant protein is used to show the full size.

between the first and second replicates with a near total loss of the Clade 1 PPTase full sized band in the second experiment as well as the appearance of several novel partial bands for the Clade 3 PPTase. This lower band of approximately 21 kD corresponds to the portion of the PPTase following the final conserved helix (Figure 2-6) as described in (Beld et al., 2014).

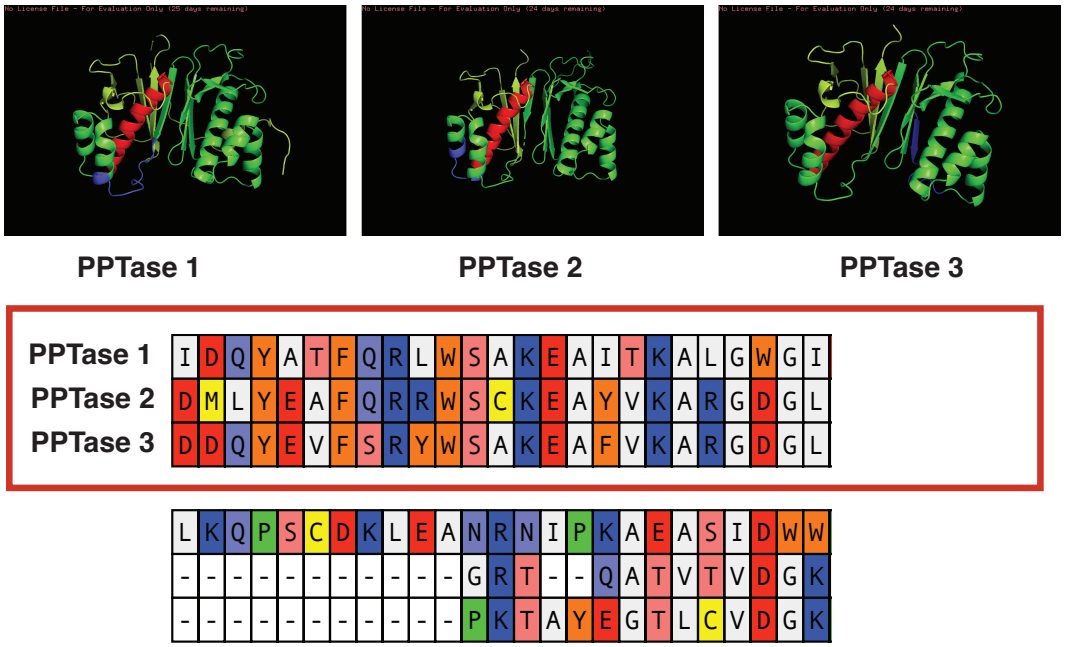



**Figure 2-6: C-terminal alignment of the theoretical cleaved portion of *Amphidinium carterae* phosphopantetheinyl transferases.**
The three *Amphidinium carterae* PPTases are shown at the top with purple indicating the anitbody epitope, red indicating the helix expected to be retained, and yellow indicating the beta sheet following the expected cleavage site based on the size of the lower band in western blots of *A. carterae* cultures. The alignment below starts with the conserved helix sequence marked with a red box followed by a short disordered region and then the beginning of the beta sheet.

Cell counts for the growth curve demonstrate sampling at the transition from lag to log phase at day 0 through log phase until day 16 and into stationary phase (Figure 2-7). The cell counts ranged from approximately 30,000 cells/ml to 145,000

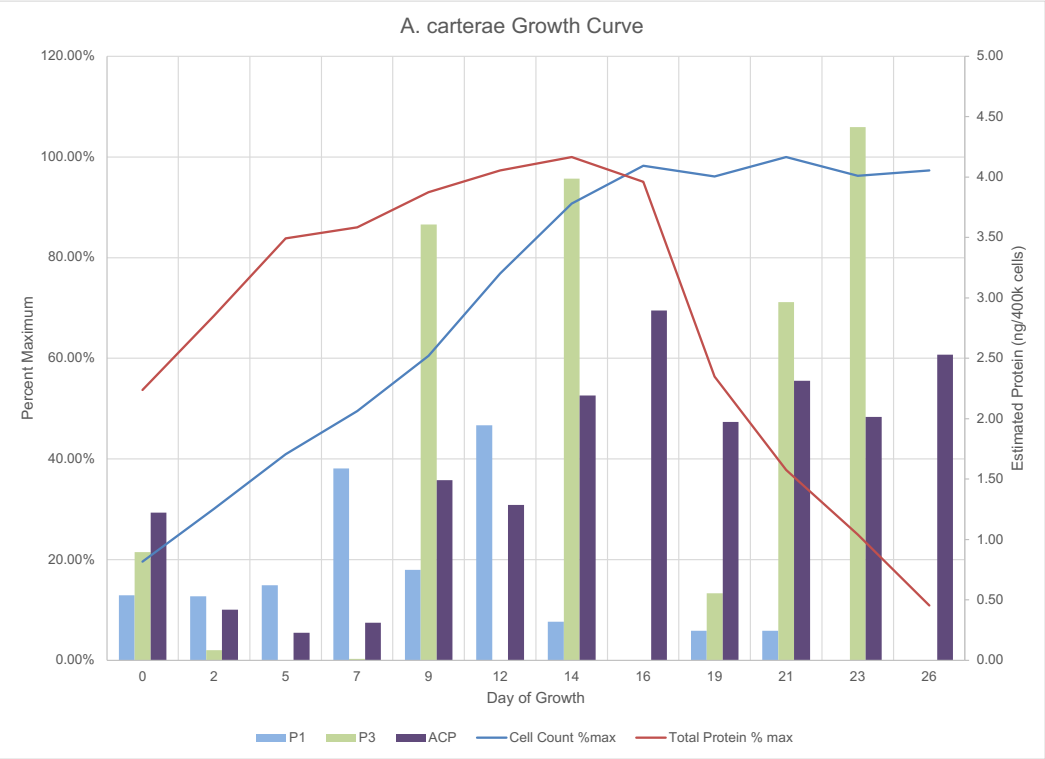



**Figure 2-7: *Amphidinium carterae* growth curve cell counts and protein quantities**
The graph shows the growth and protein measurements for an axenic *Amphidinium carterae* culture with CO2 addition. The cell counts (blue line) and total protein (red line) are shown as a percent of maximum on the left Y-axis while the western blot quantifications for the acyl carrier protein (ACP, purple) as well as the clade one and three PPTases (P1 blue, P3 green, respectively) are shown on the right Y-xaxis as an estimate of protein in ng/400k cells. The day of growth for each sample is shown on the X-axis

cells/ml with pH maintained at 7.8 ∓ 0.2. Total protein per 400,000 cell aliquot rose throughout log phase but then dropped upon entry into stationary phase. The acyl carrier protein expression seemed to have an opposite trend with a reduction upon entry into log phase, followed by an increase prior to entry into stationary phase and then a plateau. The Clade 1 and 3 PPTases seemed to have opposing expression Clade 3 expression giving way to stable Clade 1 expression, followed by an absence of Clade 1 and high expression of Clade 3 on day 14, and finally a short period of Clade 1 expression followed by Clade 3 expression. There were six time points where both the Clade 1 and 3 PPTases were expressed: days 0, 2, 9, 14, 19, and 21; and on day 16 neither were expressed. The Clade 2 PPTase was again never observed in its whole form and the same breakdown products were observed as during the diel growth experiments (Figure 2-8).

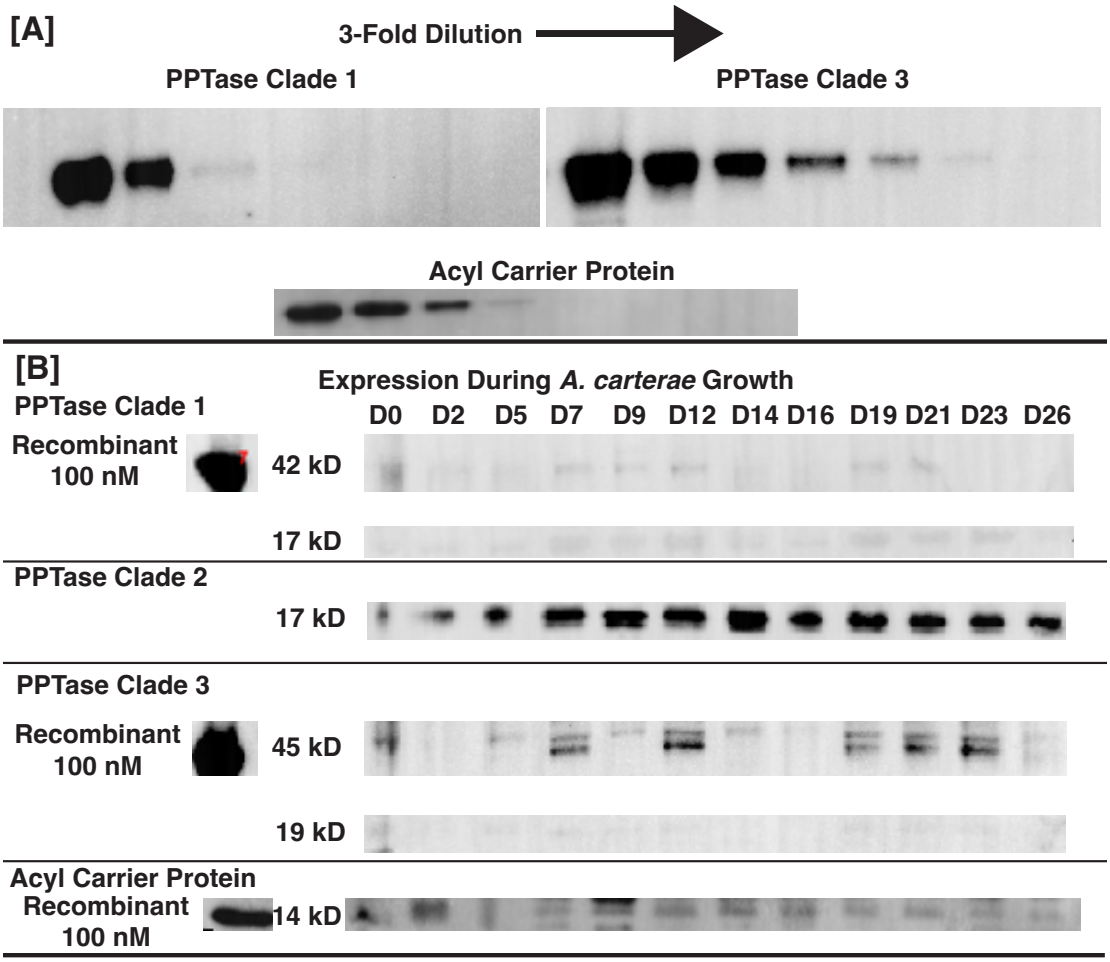



**Figure 2-8: *Amphidinium carterae* protein quantities during growth**
The upper pane (A) shows the 3-fold dilutions of the clade one and three phosphopantetheinyl transferases (PPTases) as well as the acyl carrier protein starting at 100 nM. The lower pane (B) shows the protein quantities of the clade one, two, and three PPTases as well as the acyl carrier protein. Each of the timepoints is labeled at the top of the pane with a "D" prefix for the day following CO2 control of pH. The left band is the 100 nM recombinant protein followed by the size of the band and finally the band images.

Phosphopantetheinyl transferase and acyl carrier protein sequence analysis

The 5′ and 3′ ends of all three *Amphidinium carterae* PPTases were successfully sequenced with evidence of the spliced leader in all sequences. the Clade 2 PPTase does not appear to have a stop codon while Clades 1 and 3 have a canonical dinoflagellate open reading frame (Genbank accession numbers ON157050-ON157050). The 5′ ends of all three PPTases are similar in length (51, 50 and 56 bases for Clade 1, 2 and 3, respectively), but the 3′ ends are different between Clade 1 and 2 compared to 3 with Clade 3 having the longest 3′ untranslated region (UTR). The ten bases before and after the Clade 1 PPTase stop codon are 86% similar to the same position in the Clade 2 sequence, and if a stop codon is assumed in the Clade 2

PPTase the resultant 3′ UTR is very similar in folding structure to the Clade 1 UTR compared to the Clade 3 UTR, although with greater free energy (Figure 2-9).

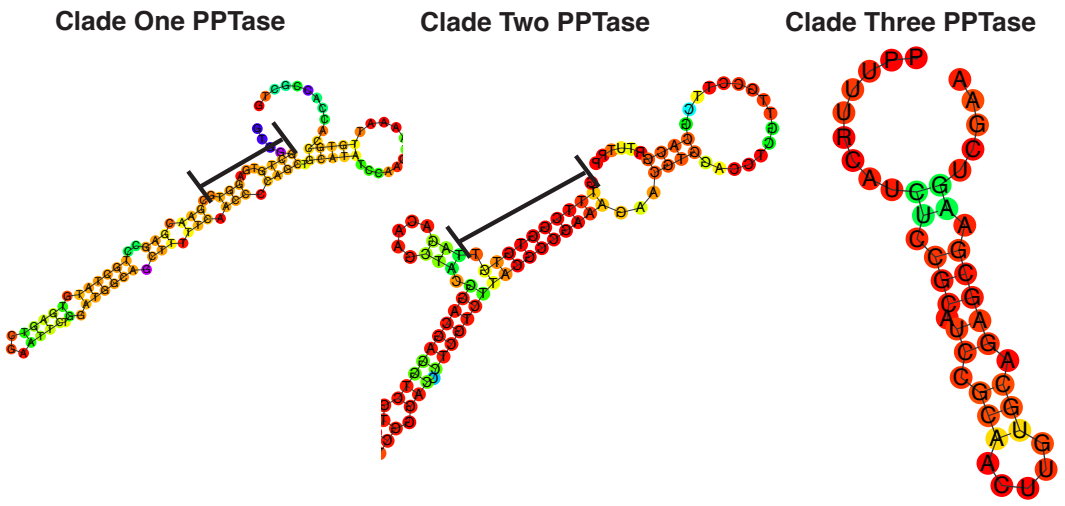



**Figure 2-9: Folding structure of the *Amphidinium carterae* phosphopantetheinyl transferase 3′ untranslated regions**
The 2-dimensionally rendered folding structure of the 3′ untranslated regions from the *Amphidinium carterae* phosphopantetheinyl transaferases are shown. The colors indicaten the relative free energy of each nucleotide from low in blue to high in red. Stem regions rich in guanidine and thymidine nucleotides are highlighted with a black bar.

SignalP as well as WolF PSORT did not detect localization signals for the three PPTases except for a weak cytoplasmic signal for the Clade 3 sequence whereas the acyl carrier protein sequence used here (ACP) has a strong chloroplast target sequence (Table 2-4). Ubiquitination sites were evident for all three PPTases

| Gene | SignalP mean | WolfP | Ubiquitination | Stability |
|---|---|---|---|---|
| PPTase 1 | 0.1112 | cytoplasm 4, plastid 3 | 0.55, 0.51 | 60.73, unstable |
| PPTase 2 | 0.131 | cytoplasm 6, plastid 4 | 0.52, 0.53, 0.65, 0.56 | 52.67, unstable |
| PPTase 3 | 0.1112 | cytoplasm 8, plastid 4 | 0.60, 0.53, 0.70, 0.59 | 38.05, stable |
| ACP | 0.782 | cytoplasm 3, plastid 10 | not analyzed | not analyzed |

with the clade 3 PPTase yielding the highest score and with multiple sites detected on every PPTase. The clade 3 PPTase was also the only protein predicted to be stable.

## *Discussion*

### The three dinoflagellate clades

The dinoflagellate phosphopantetheinyl transferases (PPTases) uncovered in this study are not similar to any of the outgroup species (Figure 2-2). The idea that lipid and natural product synthesis are easily differentiated on a sequence basis in all of life is somewhat biased considering that a small number of prokaryotes and eukaryotes have been studied. Fungi and a few Orders of bacteria have received the lion's share of attention in terms of natural product research because they make many of the compounds that we recognize and use (Jensen, 2016). Even though the underlying chemistries are similar with regards to natural product synthesis and the activation of carrier domains by PPTases, the fact is that protists are different (Sonnenschein et al., 2016; Zhu et al., 2004). What's especially intriguing with dinoflagellates are the seemingly random losses of PPTases throughout their evolution. It's difficult to tell if the roles of the PPTases are truly overlapping or if novel functionality has been acquired in certain lineages resulting in the loss of now redundant copies. *Protoceratium reticulatum* only has the Clade 3 copy and is obviously able to make lipids but additionally has a full complement of the genes to make other natural products (Williams et al., 2021). Is this a unique example or the result of the biology common to all dinoflagellates?

Many species have multiple copies within a clade (Supplementary file 2-2) indicating that the PPTases themselves can be duplicated and retained as with many other dinoflagellate genes (Bachvaroff & Place, 2008; Shoguchi et al., 2013; Stephens et al., 2020). *Oxyrrhis marina* has a single copy and is outside the dinoflagellate clade hinting at the idea that the common ancestor of core dinoflagellates only had one copy. *O. marina* is in many ways quite strange and may not represent core dinoflagellates very well (Montagnes et al., 2011). Still, it begs the question, if syndinian dinoflagellates don't have any PPTases then where did the core dinoflagellates get theirs? Horizontal gene transfer is one option and there are many examples of horizontally transferred gens in dinoflagellates (Wisecaver et al., 2013). This gene may have been acquired from the environment, but a likely suspect is the chloroplast. Transfer of genes from a plastid to the nucleus is a common means of horizontal gene transfer in dinoflagellates as well as in other stramenopiles and alveolates, whether or not the plastid has been retained during evolution (Hehenberger et al., 2019; Janouskovec et al., 2010; Keeling, 2010; Nosenko et al., 2006). If PPTases were acquired simultaneously with lipid and natural product synthesis then PPTase evolution in dinoflagellates may be on an entirely different trajectory from model bacteria and fungi where lipid synthesis was already present and natural product synthesis evolved separately.

### The biology of phosphopantetheinyl transferases in *Amphidinium carterae*

A quick comparison of PPTase transcript and protein abundance reaffirms the observation that dinoflagellates regulate gene expression post-transcriptionally with

little correlation between transcript and protein abundance (Figs. 2-3, 2-4) (Fagan et al., 1999; Lidie et al., 2005; Morse et al., 1989). The dip in transcript abundance just after lights out likely correlates with the synthesis of new DNA just prior to cell division that would inhibit the transcription of new RNA. Strikingly, there are major portions of this same time period where the Clade 3 PPTase protein is not observed meaning that a transcript is made without a protein product expressed. This does not appear to be degradation since neither the full form nor the partial form are visible. The same is true of the Clade 1 PPTase that is not present in its full form at the early time points but the breakdown product is visible (Figure 2-5). It was in an effort to prove this disappearance that the diel experiment was repeated. At first glance the diel expression in the initial experiment seems reasonable with the Clade 3 PPTase lowering in expression near the end of the day while the Clade 1 PPTase peaks at lights out indicating that one could participate in natural product synthesis while the other performs lipid synthesis, mediated by the circadian cycle. The replicated experiment showed a very different result with almost no evidence for the full form of either PPTase at any time point (Figure 2-5). There was also evidence for sporadic or low-level expression of the Clade 3 PPTase with a few lower bands visible at some time points. The regulation of these PPTases appears to be dynamic and even linked. During the growth curve the Clade 1 and 3 PPTases again show an alternating expression pattern with one increasing while the other decreases (Figure 2-6) pointing to a feedback loop with the two PPTases influencing each other's expression. The three PPTases also appear to have a common mechanism for inactivation via the cleavage of the C-terminus resulting in the observed lower band (Figure 2-7). The expression of the PPTases seems to be totally unrelated to acyl carrier protein expression at first glance since the acyl carrier protein increases steadily in expression during log phase and then plateaus during the entrance into stationary phase. When the acyl carrier protein expression is maximal at day 16 neither of the PPTases are observed. The acyl carrier protein could be subtly informing the expression of the PPTases since in other systems the acyl carrier protein is always observed in the phosphopantetheinated form (Flugel, Hwangbo, Lambalot, Cronan, & Walsh, 2000) meaning that the total acyl carrier protein expression isn't what's important but rather the portion that isn't phosphopantetheinated. Either way, the expression patterns of the PPTases appear to be quite dynamic but do not appear to be related to the synthesis of lipids or natural products in any specific manner.

The Clade 2 PPTase is quite strange with a full protein not observed in any portion of this study (Figs. 2-5,8). This is likely due to the absence of a stop codon but it still doesn't explain why this protein is expressed at all. Both the Clade 1 and 2 PPTases appear to have similar stem loop structures in the 3′ UTR with stretches of guanine and thymine residues in the stem that have been previously associated with circadian expression patterns (Figure 2-9) (Fagan et al., 1999; Lapointe & Morse, 2008). While the expression of the Clade 1 PPTase is obviously more complicated than simple circadian expression, the multitude of circadian expressed genes in dinoflagellates (Akimoto, Wu, Kinumi, & Ohmiya, 2004) and the expression of the seemingly non-functional Clade 2 PPTase raises the question of the biological role of the constant expression and degradation of dinoflagellate proteins. An idea proposed by Woody Hastings was that this is a way of recycling nitrogen (Hastings, 2013). The

fact that constant protein recycling is observed in a critical gene such as a PPTase demonstrates how widespread this biological phenomenon may be in dinoflagellates and points to the importance of looking at protein expression over multiple time points when characterizing dinoflagellate genes.

## *Conclusion*

The phosphopantetheinyl transferases of dinoflagellates are quite atypical with multiple gains and losses through their evolutionary history that do not correlate with a biological process. Likewise, the expression patterns of either the Clade 1 or 3 PPTases do not correlate with the expression of the acyl carrier protein furthering the notion that the functional segregation of PPTases that has been the canon in bacteria and fungi does not apply in dinoflagellates. In some ways dinoflagellates are unique with regards to how often horizontal gene transfer is observed. Thus, it may not be surprising that the evolution of PPTases is dinoflagellates breaks the norms. In order to investigate this further the PPTases themselves need to be assessed for their function with regards to the acyl carrier protein. If there is a functional constraint, then one could expect one or more PPTase to be selectively functional for this essential carrier protein. Also, a broader survey into the PPTases of other photosynthetic algae, especially the likely origins of dinoflagellate chloroplasts may shed light on how these enzymes function in the broader evolutionary sense. After all, fungi are quite distantly related to dinoflagellates and there isn't a very good framework for protist biology to make these kinds of comparisons. This work will hopefully contribute to our understanding of dinoflagellate PPTases by demonstrating just how different they are from other organisms.

# Chapter 3: In-vivo and *In vitro* Binding Assays with Dinoflagellate Thiolation Domains


## *Abstract*

Photosynthetic dinoflagellates synthesize many toxic but also potential therapeutic compounds therapeutics via polyketide/non-ribosomal peptide synthesis, a common means of producing natural products in bacteria and fungi. Although canonical genes are identifiable in dinoflagellate transcriptomes, the biosynthetic pathways are obfuscated by high copy numbers and fractured synteny. This study focuses on the carrier domains that scaffold natural product synthesis (thiolation domains) and the phosphopantetheinyl transferases (PPTases) that thiolate these carriers. We replaced the thiolation domain of the indigoidine producing BpsA gene from the bacterium *Streptomyces lavendulae* with those of three multidomain dinoflagellate transcripts and coexpressed these constructs with each of three dinoflagellate PPTases looking for specific pairings that would identify distinct pathways. These protein products were also purified from *E. coli* or synthesized to perform the same assays in vitro. Successful interactions were measured indirectly by the production of indigoidine or indirectly by quantifying the amount of phosphopantetheinate added to the thiolation domain by each transferase. Unsurprisingly, several of the dinoflagellate thiolation domains when incorporated reduced or removed the ability of the bacterial reporter to synthesize indigoidine despite being successfully phosphopantetheinated. What was surprising was that all the transferases were able to phosphopantetheinate all the thiolation domains nearly equally, defying the canon that each transferase is specific for a single process via binding specificity. The broad substrate recognition shown here help explain why phosphopantetheinyl transferases are lost throughout dinoflagellate evolution without a loss in a biochemical process but also how new thiolation domains seem to be acquired through horizontal gene transfer and retained in evolutionary lineages. It is also hoped that the techniques presented here will be used to validate other functional assignments where gene copy number is an issue.


## *Introduction*

Dinoflagellates make a variety of natural products that have largely been identified based on their impact to human and animal health (Deeds, Terlizzi, Adolf, Stoecker, & Place, 2002; Twiner et al., 2012; Walsh et al., 2015; Wang, 2008). The actual biological and/or ecological roles are largely unknown and require further study. The exceptions include karlotoxin, the only toxin known to be actively released

from the cell for prey capture and predator avoidance (Adolf et al., 2007; Sheng et al., 2010), and brevetoxin that likely functions as an indicator of redox state in the chloroplast (Chen et al., 2018; Colon et al., 2021). This functional knowledge gap is exacerbated by a lack of a biosynthetic framework that would allow a more thorough cataloging of the natural products produced by dinoflagellates as well as insights into their evolution.

Natural product synthesis has been extensively studied in bacteria and fungi yielding a mechanistic framework that operates as a series of modules with repeated chemistries followed by some modifications resulting in the final molecule. Essentially, small carboxylic acids are added to the thiol end of a phosphopantetheinate group attached to the serine of a carrier protein (Beld et al., 2014) via a condensation reaction that releases either carbon dioxide or water with prior activation by ATP (Bentley & Bennett, 1999; Khosla, 2009; Sieber & Marahiel, 2005). These building blocks are then modified by subsequent reduction, methylation, carbon deletion, and other rarer reactions before the next carboxylic acid is added. In general these are added by genetic modules comprised of single proteins with multiple functional domains or multiple cis-acting proteins brought together to form an enzymatic complex, although trans-acting elements are not uncommon (Khosla et al., 2009; Rausch et al., 2007; Wang et al., 2014) and substrates from multiple pathways can be combined (Franke et al., 2012; Kevany et al., 2009).

Research into the biosynthesis of many natural products has relied heavily on the fact that gene arrangement is strongly predictive of a given natural product's final structure. Unfortunately, dinoflagellate genomes are large and heavily duplicated (Bachvaroff & Place, 2008), although mass spectrometry and NMR have been able to readily identify that dinoflagellate toxins have the hallmarks of classic natural product synthesis (Fukatsu et al., 2007; Ishida et al., 1995; Meng et al., 2010; Peng et al., 2010; Sasaki et al., 1996; Satake et al., 1997; Seki et al., 1995; Van Wagoner et al., 2008; Van Wagoner et al., 2010; Wright et al., 1996), with some exceptions (Van Wagoner et al., 2014). Investigations into genes potentially involved in toxin synthesis have had some success (Beedessee et al., 2020; Snyder et al., 2003; Verma et al., 2019), most notably in the separation of genes involved in natural product synthesis from the analogous synthesis of lipids (Kohli et al., 2015; Kohli et al., 2016; Meyer et al., 2015; Van Dolah et al., 2017) and the identification of multi-domain genes (Bachvaroff, Williams, Jagus, & Place, 2015; Kohli et al., 2017; Van Dolah et al., 2020). These multi-domain genes can then be used to further bin single domain genes into functional groups, although from here the waters get quite muddy with uneven gene copy numbers and the unprecedented duplication of genes related to lipid synthesis (Williams et al., 2021). Thus, in many ways sequence analysis has reached its limits in its ability to shed light on the synthesis of dinoflagellate natural products.

The aim of this project is to extend the current sequence-based knowledge into a biochemical based understanding of natural product synthesis by expressing dinoflagellate proteins in a heterologous system. An attractive target is the carrier protein called the thiolation domain that is activated by the attachment of the phosphopantetheinate group of Coenzyme A by a phosphopantetheinyl transferase (PPTase) creating a free thiol moiety. This is the first rate limiting step in natural

product synthesis and provides the substrate upon which the actual anabolism is performed by all of the catalytic enzymes. Generally, the activation of a thiolation domain by any phosphopantetheinyl transferase is highly specific and separates specific biosynthetic pathways, although the actual transfer of a phosphopantetheinyl group is not required for recognition of the transferase to a thiolation domain (Bunkoczi et al., 2007). The thiolation domains of dinoflagellates can be readily separated into two main groups indicative of lipids and natural products (Williams et al., 2021). Although the number of thiolation domains can number above one-hundred, the number of phosphopantetheinyl transferase activators is no more than three (Williams, Bachvaroff, & Place, 2020). Thiolation domain activation will be tracked using the blue pigment synthesizing agent (BpsA gene) from *Streptomyces lavendulae* (Takahashi et al., 2007). This single gene contains several domains, including the essential thiolation domain, that coordinately produce the blue dye indigoidine. It has been previously used to differentiate lipid type and natural product type PPTases because its thiolation domain is naturally recognized by PPTases that act in natural product synthesis (Owen, Copp, & Ackerley, 2011). The rationale is that if a given dinoflagellate PPTase can activate the thiolation domain of the BpsA reporter then indigoidine will be produced. This pairing of activator and thiolation domain is a common method for determining specificity (Geerlof, Lewendon, & Shaw, 1999; Murugan, Kong, Sun, Rao, & Liang, 2010) and has been performed in some protists with a surprising promiscuity not found in bacteria and fungi (Cai, Herschap, & Zhu, 2005; Sonnenschein et al., 2016). This study advances previous work by replacing the thiolation domain of the BpsA reporter with several different dinoflagellate sequences to allow for the pairing of each activator with a multitude of potential phosphopantetheination sites. Although the integration of dinoflagellate sequence into the bacterially derived reporter was largely successful, some of the thiolation domains disrupted the activity of the BpsA gene necessitating additional assays that could directly detect phosphopantetheination. Also, the acyl carrier protein could not be expressed in *E. coli* and instead was synthesized in vitro and assessed directly for phosphopantetheination. All of the thiolation domains used, including the acyl carrier protein could be phosphopantetheinated by all the PPTases, calling into question the canon that PPTases are the gatekeepers for different pathways in dinoflagellates.

## *Materials and Methods*

### BpsA reporter modification and use

The BpsA reporter described in Owen *et al.* 2011 was kindly obtained from the Ackerley lab at the University of Victoria in Wellington New Zealand. The region encompassing the thiolation domain was amplified with the primers “BpsA_outF2” and “BpsA_outR2” listed in Table 1 at 500 nm final concentration and 10 µg of vector template using the Phusion high fidelity polymerase (New England Biolabs,

Cambridge MA) as follows: Initial denaturation at 98 °C for 2′; followed by 40 cycles of denaturation at 95 °C for 15 seconds, annealing at 58 °C for 20″, and extension at 72 °C for 1′ 30″; and polishing at 72 °C for 5′. The amplified product, termed "BpsA_insert0" was purified and sequenced at the BioAnalytical Services Laboratory (BASlab) at the Institute for Marine and Environmental Science on an Applied Biosystems 3130 XL. This sequence was used to design the remaining primers in Table 1 to insert a HindIII site in the 3′ end of the thiolation domain and an AflII site in the 5′ end as described in the primer name. The insertions result in the shift of arginine to a lysine at the HindIII site.

The HindIII and AflII sites were incorporated into the vector in a two-stage process. First the HindIII site was created via two amplifications using "BpsA_outF2" with "BpsA_hindiiiR2" and "BpsA_outR2" with "BpsA_hindiiiF2" with the same reaction conditions as the thiolation domain amplification. The resultant products were purified using a DNA Clean and Concentrate-5 kit from Zymo research (Irvine CA) and eluted into 10ul of distilled deionized water. Approximately 2.5 µg or product was digested with the HindIII-HF restriction enzyme from New England Biolabs for four hours at 37 °C, separated on an ethidium bromide impregnated 1% agarose gel in 0.5X TBE at 15 V/cm for 50 minutes, excised under ultraviolet illumination, and purified using a Monarch DNA Gel Extraction kit from New England Biolabs as directed. The two digested fragments were then combined and ligated using a T4 ligase from Promega (Hercules CA) overnight at 18 °C. This product was then used as template for the second stage amplification using primers "BpsA_outF2" with "BpsA_afliiR2" and "BpsA_outR2" with "BspA_afliiF2" using the same conditions as the HindIII site amplification. This was purified, digested with AflII restriction enzyme from New England Biolabs, agarose gel purified, and combined and ligated in the same manner as the HindIII products resulting in "BpsA_insert1" (Supplementary File 3-1, Figure 3-1).

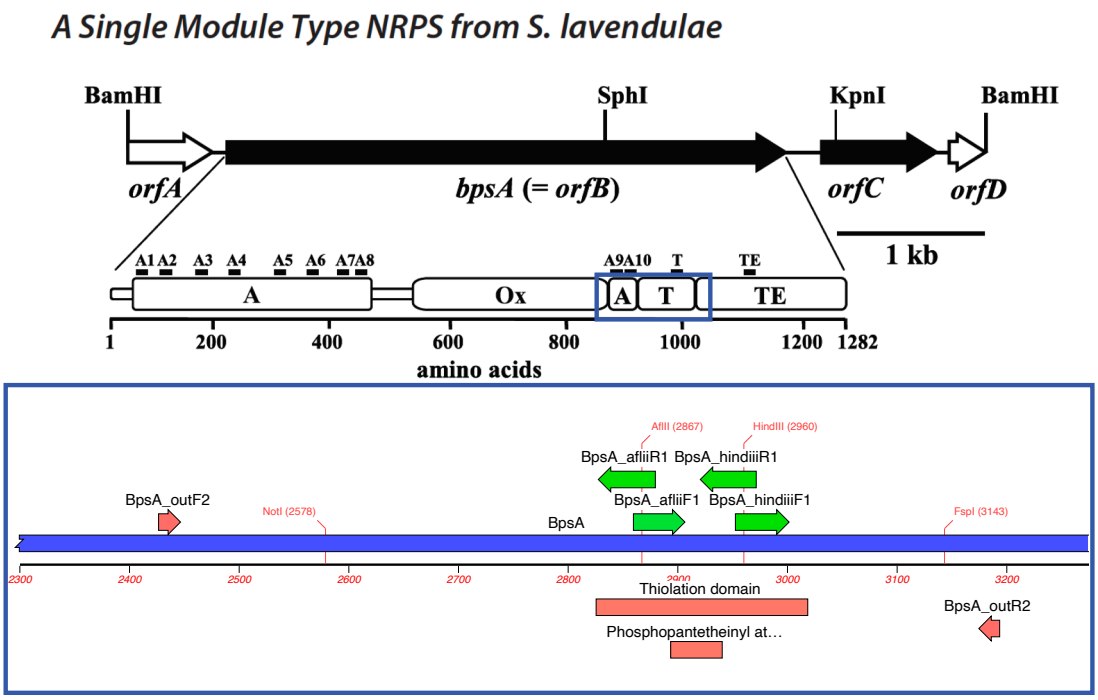



**Figure 3-1: A modification of the BpsA to allow the insertion of dinoflagellate thiolation domain sequences.**
Shown above is the BpsA gene from *Streptomyces lavendulae* originally published in Takahashi *et al.* 2007 showing each of the domains with the thiolation domain marked with a "T". The region surrounded by a blue box is expanded in the bottom showing the thiolation domain and the phosphopantetheinate transferase binding site as red boxes. The existing NotI and FspI restriction as well as the introduced AflII and HindIII sites are shown in red text. The primers used to islolate this region and attach the novel restriction sites are shown as green arrows above with the primer direction indicated by the arrow direction.

BpsA_insert1 was amplified using the same conditions as the original thiolation domain and purified using the DNA Clean and Concentrate-5 kit. This product as well as the original BpsA vector were double digested with the NotI-HF and FspI restriction enzymes from New England Biolabs at 37 °C overnight in cutsmart buffer followed by agarose gel purification and ligation as with the HindIII and AflII amplicons resulting in the BpsA2.1 vector. This was amplified using the Templiphi 100 kit from Cytiva and cloned into *E. coli* JM109 from Promega according to the manufacturer's directions. A selection of the resultant colonies was grown and the plasmid extracted for co-expression in BL21(DE3) cells with each of the PPTases from *Amphidinium carterae* as well as PcpS from *Pseudomonas aeruginosa* as a positive control to confirm the retained activity of the modified BpsA reporter (Figure 3-2).

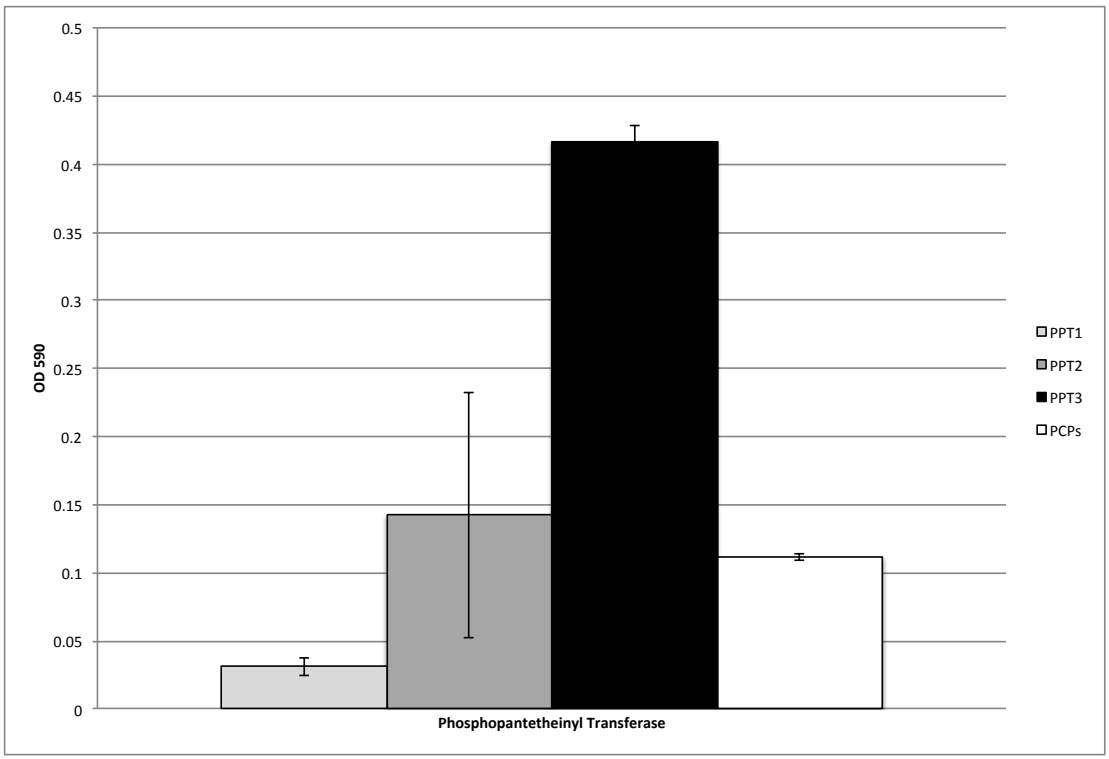



**Figure 3-2: Coexpression of PPTases with the BpsA reporter.**
Indigoidine production is shown as an absorbance at 590 nm on the Y-axis for the coexpression of each of the three *Amphidinium carterae* as well as PcpS from *Pseudomonas aeruginosa* with the BpsA reporter. Error bars represent standard deviation from triplicate liquid cultures.

Insertion of dinoflagellate thiolation domains and co-expression in *E. coli*

The natural product associated thiolation domains (Williams et al., 2021) in three multi-domain transcripts (Figure 3-3) (Bachvaroff, Williams, Jagus, & Place, 2015; Kohli et al., 2017; Van Dolah et al., 2017) from *A. carterae* were chosen for

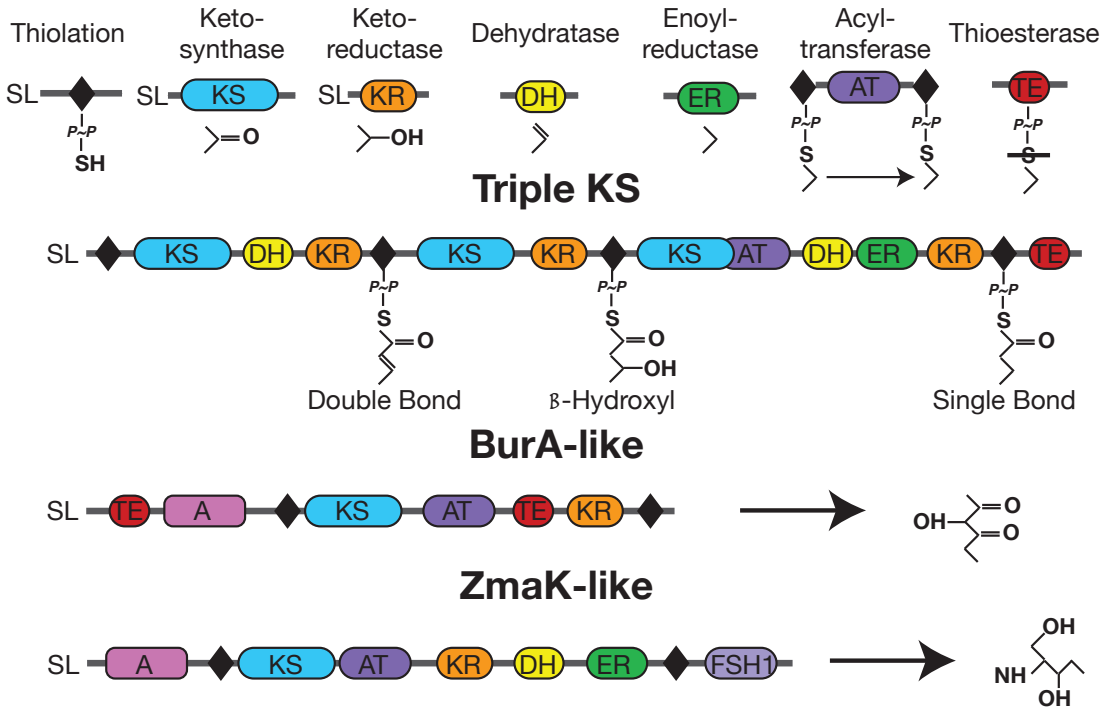



**Figure 3-3: Domain arrangement of *A. carterae* transcripts containing thiolation domains used in this study.**
Individual modular synthase domains are shown at the top with example products for their reaction. In addition Adenylation (A) and FSH1 serine hydrolases (FSH1) are shown for the multi-domain transcripts with examples of potential products included. The phosphopantetheinate group is shown as "P~P" with a single bind to a sulfur. "SL" refers to the dinoflagellate spliced leader sequence and is present if a spliced leader sequence has been verified.

complementation in *E. coli* with the three *A. carterae* phosphopantetheinyl transferases (PPTases) that could activate them (Figure 3-4). These were termed

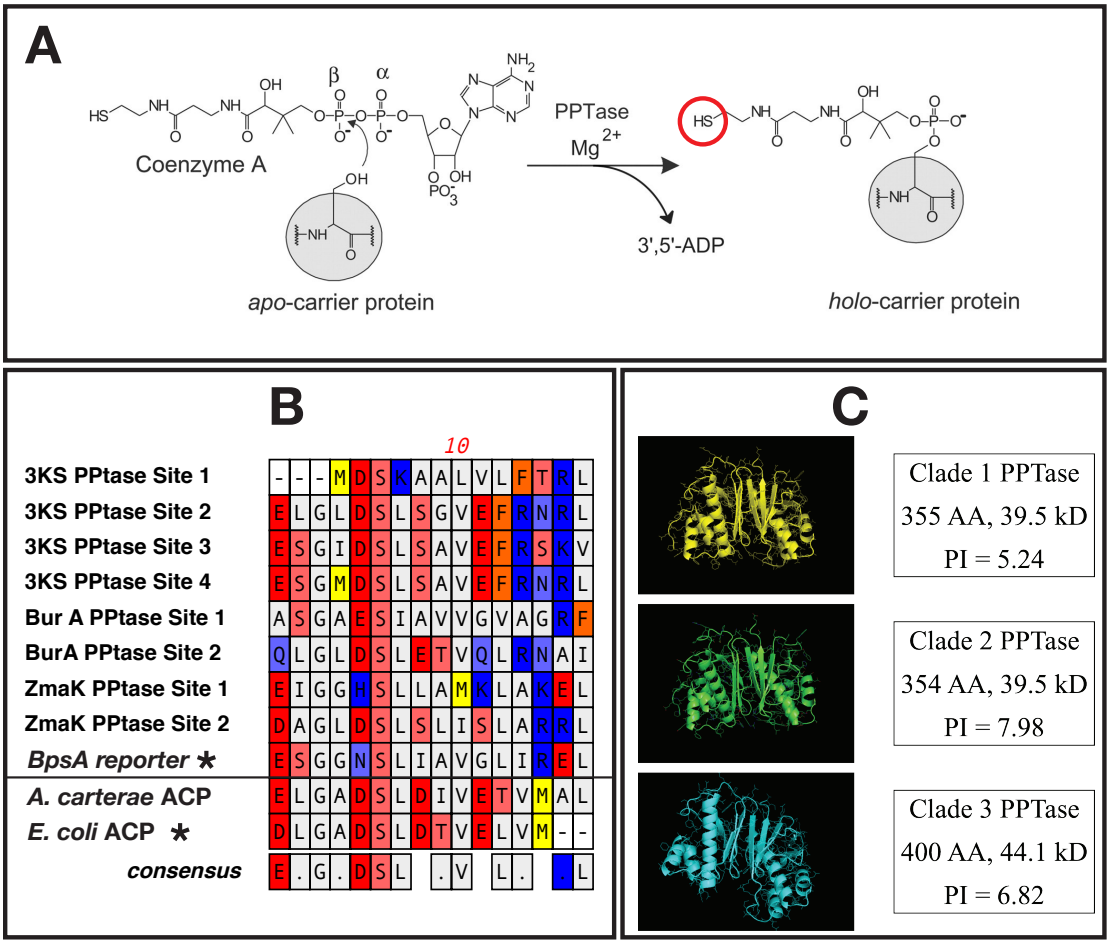



**Figure 3-4: A mechanism of phosphopantetheination and the dinoflagellate thiolation domains used in this study.**
**A**) A diagram of the phosphopantetheination reaction from Finking et al. 2002 showing the phosphopantetheinate arm of coenzyme A attachment to the serine of a carrier protei or domain resulting in a free thiol group (red circle). **B**) The amino acid sequences of the thiolation domains from *A. carterae* used in this study except those marked with a "*" that are from the *S. lavendulae* isolated BpsA gene and the acyl carrier protein (ACP) from *E. coli*. Sequences above the line are theorized to be involved in natural product synthesis while those below the line are for lipid synthesis. **C**) The predicted foiding for the three phosphopantetheinate transferases from *A. carterae*.

"3KS" for the three ketosynthase domains present, "BurA" for its similarity in sequence and domain arrangement to the BurA gene in *Burkholderia species* (Franke et al., 2012), and "ZmaK" for the sequence similarity of the dinoflagellate adenylation domain in this transcript to the *Bacillus cereus* adenylation domain in the ZmaK cluster (Kevany et al., 2009). Each individual thiolation domain was named according to the transcript it was derived from followed by a numeral indicating the order from 5′ to 3′ in the transcript, *e.g.,* "3KS3" would be the third thiolation domain in the three ketosynthase domain containing transcript. The PPTase binding site amino acid sequence (Table 3-1) of each thiolation domain was codon optimized

| Primers | | | |
|---|---|---|---|
| **Primer Name** | **Sequence 5′:3′** | **Length** | **Annealing ° C** |
| BpsA_outF2 | TCCAGCACCTGATGATGAAC | 20 | 58.4 |
| BpsA_outR2 | CTGGATGCCGTAGAACGAG | 19 | 59.5 |
| BpsAhindiiiR1 | GACGCCAAGCTTCGCGTTGAGCTCGCGGAC GAGGCCGACGGCGATCAGCGA | 51 | 91.1 |
| BpsAhindiiiF1 | CAACGCGAAGCTTGGCGTCTCCCTGCCGCTG CAGAGCGTCCTGGAGTCC | 49 | 89.6 |
| BpsAafliiR1 | CTCGCGCTTAAGGGCCTTCTCCCAGACCGCC GCGATCTCCTTCTCCGT | 48 | 88.5 |
| BpsAafliiF1 | AGAAGGCCCTTAAGCGCGAGAACGCCTCCGT CCAGGACGACTTCTTCG | 48 | 86.4 |
| **Inserts** | | | |
| **Insert Name** | **Sequence 5′:3′** | **Binding Site Amino Acid** | |
| 3KS_1 | GAATCGGGCATGGACTCAAAAGCAGCCCTTGT TCTG | ***ESG*** MDSKAALVL | |
| 3KS_2 | GAATTGGGCTTAGATTCTTTGTCCGGCGTTGA ATTT | ELGLDSLSGVEF | |
| 3KS_3 | GAAAGCGGAATTGATTCCTTGTCTGCAGTAGA GTTT | ESGIDSLSAVEF | |
| 3KS_4 | GAGAGTGGCATGGACTCATTATCTGCCGTCGA GTTT | ESGMDSLSAVEF | |
| BurA_1 | GCT TCA GGT GCA GAA TCT ATC GCT GTC GTG GGC GTG | ASGAESIAVVGV | |
| BurA_2 | CAA TTA GGA TTA GAC AGC TTG GAA ACC GTT CAA CTG | QLGLDSLETVQL | |
| ZmaK_1 | GAA ATC GGT GGG CAC TCG CTG TTA GCA ATG AAA CTT | EIGGHSLLAMKL | |
| ZmaK_2 | GAT GCC GGG TTA GAT AGC TTA TCC TTA ATT AGC TTA | DAGLDSLSLISL | |
| 5′ Linker † | AGAAGGCCCTTAAGCGCGAGAACGCCTCCGTCCAGGACGACTTCTTC | | |
| 3′ Linker † | GTCCGCGAGCTCAACGCGAAGCTTGGCGTCTCCCTGCCGCTG | | |
| **"†" denotes common linkers to all other inserts and were placed at the 5 and 3′ ends during synthesis as indicated** | | | |
| **The "3KS_1" sequence shown in bolded italics is the wild type sequeence included to ensure consistent insert size** | | | |

for expression in *E. coli* and ordered as an oligonucleotide from Integrated DNA Technologies. Each oligonucleotide was synthesized with common linker sequences containing the AflII and HindIII restriction sites in the BpsA2.1 plasmid, one for the 5′ end and one for the 3′ end (Table 3-1). Thus, each oligonucleotide consisted of the 5′ linker followed by the unique thiolation domain sequence and then the 3′ linker.

For each thiolation domain, the synthetic oligonucleotide as well as the BpsA2.1 plasmid were double digested with HindIII and AflII overnight at 37 °C in cutsmart buffer followed by agarose gel purification using a Monarch DNA Gel

Extraction kit from New England Biolabs. The cut insert and plasmid were combined and ligated with a T4 ligase (Promega) at 18 °C overnight. Each ligated plasmid was amplified with the Templiphi 100 kit from Cytiva and cloned into *E. coli* JM109 from Promega according to the manufacturer's directions. JM109 clones were sequenced to verify the presence of the dinoflagellate insert in the plasmid followed by alkaline extraction (Sambrook, Fritsch, & Maniatis, 1989). Plasmids were then cloned into chemically competent BL21(DE3) *E. coli* (Thermo Fisher) along with one of the three PPTase activators (Figure 3-4) from *A. carterae* in a separate pET-20b plasmid according to the directions for the competent cells and plated onto LB agar containing 100 µg/ml carbenicillin and 50 µg/ml spectinomycin at 37 °C. Additionally, each PPTase vector and the thiolation domain vectors were individually cloned into BL21(DE3) to assess protein expression. The vectors for the PPTases were chosen to have a different replication sequence than the reporter to avoid conflicts during growth. Colonies were picked, grown in liquid media containing antibiotics overnight at 37 °C and stored at -80 °C with glycerol added to a final concentration of 12% V/V. For assessment of protein expression glycerol stocks were used to inoculate 10 ml of LB in a 250ml Erlenmeyer with appropriate antibiotics and grown overnight at 37 °C with shaking at 250 rpm. This was then diluted into 500 ml of LB media with antibiotics in a 2000 ml Erlenmeyer followed by a reduction of temperature to 30 °C and growth for 3 hours with shaking. Protein expression was induced by the addition of 500 µl of 0.1M IPTG followed by incubation at 25 °C for 3 hours with shaking. Cells were spun at 10k x g for 10 minutes at 4 °C, and the media was decanted. Cells were suspended in 20 ml of PBS at 4 °C with a bacterial protease inhibitor (Sigma Aldrich, St. Louis MO), and proteins were extracted in a French press at 20k local PSI followed by centrifugation at 10k x g for 10 minutes at 4 °C to separate soluble and insoluble material. Insoluble proteins were recovered from the pellet by the addition of 6M urea in equal volume to the supernatant. Heterologous proteins were purified with a 1 ml HiTrap Talon crude column (Cytiva) on an AKTA chromatography system with elution into 50 mM Tris with 250 mM imidazole. Proteins were separated by SDS-PAGE electrophoresis with 4-12% bis-tris gels (ThermoFisher) and imaged with Imperial Coomassie stain (ThermoFisher).

*E. coli* clones containing one of the three *A. carterae* PPTases and one of the eight BpsA reporters with dinoflagellate thiolation domain sequence were each grown onto agar plates containing “autoinduction” media (Studier, 2005). Colonies were grown at 25 °C for 48 hours to allow for growth, protein expression, and indigoidine production. The plates were photographed, and each colony was assessed for dye production by measurement of grayscale density using image J (https://imagej.net/) with the space in between colonies as a baseline for background subtraction.

### Reporter and PPTase protein expression and purification

The *E. coli* BL21(DE3) cells containing pET-20b plasmids with each of the three codon optimized PPTase sequences were grown in autoinduction media (Studier, 2005) with 100 µg/ml carbenicillin at 25 ℃ for three days at 250 rpm. Likewise, the pCDFDuet-1 plasmid containing the BpsA gene with and without the dinoflagellate

thiolation domain inserts 3KS4, BurA1, ZmaK1, and ACP (Figure 3-4) were grown in *E. coli* BL21(DE3) cells. These constructs could not be expressed by autoinduction since production at temperatures above 20 ℃ did not produce active BpsA protein capable of producing indigoidine. Instead, 10 ml LB media containing 50 µg/ml spectinomycin was inoculated with each of the dinoflagellate thiolation domain containing *E. coli* clones and grown overnight at 37 ℃ at 250 rpm in a 250 ml Erlenmeyer flask. These cultures were then diluted to 500 ml LB with spectinomycin and grown to an OD 600 of 6.0. Cultures were then chilled to 18 ℃ in an ice bath, induced with 500 µl 0.1M IPTG, and grown overnight at 18 ℃ at 250 RPM.

Induced *E. coli* clones were collected after protein expression by centrifugation at 10,000 x g for 15 minutes at 4 ℃ and the supernatant was decanted. Pellets were stored at -80 ℃ until processing. Frozen pellets were suspended in 25 ml of lysis buffer containing 5 mM imidazole, 500 mM NaCl, 25 % (v/v) glycerol and 20 mM Tris/HCl pH 7.5 along with bacterial protease inhibitors (Sigma Aldrich, St. Louis MO) and thawed at 4 ℃. Cells were lysed with a French press chilled to 4 ℃ with 1000 PSI at the piston and 20,000 PSI at the outlet. The lysate was clarified by centrifugation at 12,000 x g for 15 minutes at 4 ℃ and the supernatant was treated with benzonase (Thermo Fisher) overnight at 4 ℃. The His-tagged lysate was bound to a 1 ml Hi-trap crude cobalt column (Cytiva); washed with 25 column volumes of 5mM Imidazole, 250 mM NaCl, 20 mM Tris/HCl pH 7.5; and eluted into 10 separate volumes of 250 mM Imidazole, 250 mM NaCl, 20mM Tris/HCl pH 7.5, and 12.5% glycerol using an AKTA chromatography system (Cytiva). Elution fractions were separated with 4-12% Bis-Tris gels from Novex (Thermo Fisher) in MOPS buffer at a constant voltage of 165 V for 50 minutes and stained with Imperial stain (Thermo Fisher) according to the manufacturer's directions to verify protein capture. Fractions containing expressed protein were concentrated using a 10 ml Amicon high pressure stirred filter apparatus with a 5 kD regenerated cellulose filter for the PPTases and 30 kD filters for the BpsA constructs at 45 PSI with Nitrogen at 4℃ (Sigma Aldrich). Proteins were then buffer exchanged with 50 mM TRIS pH 7.5 using an amicon 500 µl spin column with a 3 kD nominal pore size for the PPTases and 10 kD for the BpsA constructs at 4 ℃ (Sigma Aldrich). The BpsA vector containing the acyl carrier protein (ACP) thiolation domain could not be expressed in *E. coli* and the PPTase from dinoflagellate Clade 2 was retained in the insoluble pellet following lysis. For these proteins, the Pure Express kit (New England Biolabs) was used with approximately 10 ng of pET-18b plasmid containing either the PPTase Clade 2 or the entire ACP open reading frames according to the manufacturer's direction for synthesis in vitro. This means that the ACP thiolation domain was not expressed in the BpsA framework and was not used for indigoidine based detection of phosphopantetheination by the three dinoflagellate PPTases but was available for the direct measurement of free thiol. Artificial ribosomes from the Pure Express kit were removed by passage through a 100 kD Amicon spin column at 10,000 x g for 20 minutes followed by buffer exchange and concentrated using a 3 kD Amicon spin column with two washes at 500 µl each in the same manner as the purified PPTases expressed in *E. coli*.

Purified protein in 50 mM TRIS was prepared for storage by the addition of glycerol to a final concentration of 12.5% and quantified using a Q-bit protein

quantification kit (Thermo Fisher). This was adjusted based on the band densities following imperial staining to account for spurious bands from His-tag purification.

### Estimation of thiolation domain phosphopantetheination in vitro

Each of the three BpsA constructs containing the dinoflagellate thiolation domains along with the original BpsA thiolation domain were combined with each of the three dinoflagellate PPTases according to the protocols from Owen *et al*. 2011. Briefly, a premix containing one of each PPTase at 0.2 µM and one of each BpsA construct in two-fold dilutions from 8 µM to 0.25 µM in 50 mM Tris pH 7.5, 10 mM MgCl2, and 100 µM CoA was incubated at 30 ℃ for 10 minutes. This was then combined with a ⅓ volume of 25 mM glutamine and 5 mM ATP for a final volume of 50 µl. The absorbance at 590 nm was measured every 10 seconds for 30 minutes to quantify the indigoidine production using a Spectramax i3x plate reader (Molecular Devices, San Jose, CA). Windows of linear indigoidine production were used to determine the maximum rate of production for each combination of PPTase and BpsA constructs (Figure 3-5).

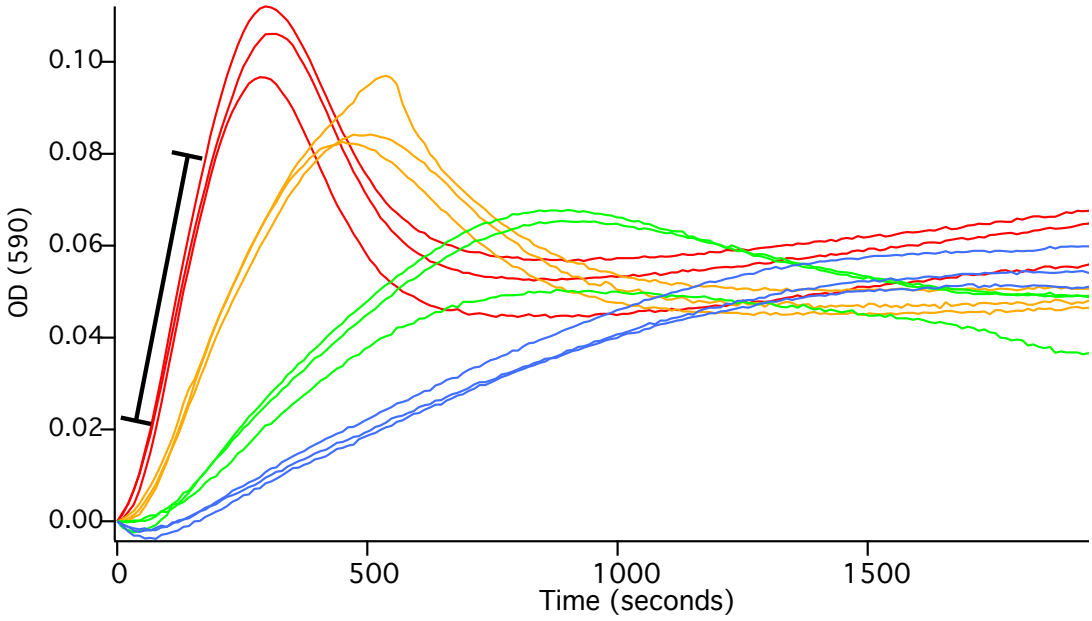



**Figure 3-5: Concentration dependent Indigoidine production.**
The kinetics of indigoidine production are shown with the change in absorbance at 590 nm on the Y-axis and the reaction time X-axis using the Clade 3 PPTase and the wild-type BpsA reporter. Triplicate reactions are shown with red lines for 8 $\mu$m, yellow for 4 $\mu$m, green for 2 $\mu$m, and blue for 1 $\mu$m BpsA reporter. An example of linear production of indigoidine used to calculate rate is shown as a black bar next to the 8 $\mu$m replicates.

The addition of phosphopantetheinate was quantified using the free thiol detection kit from Abcam (Cambridge, UK). A seven-point standard curve was

performed using 5-fold dilutions of CoA in 50 mM TRIS pH 7.5 according to the directions of the kit with 50 mM TRIS pH 7.5 as a blank control (Figure 3-6).

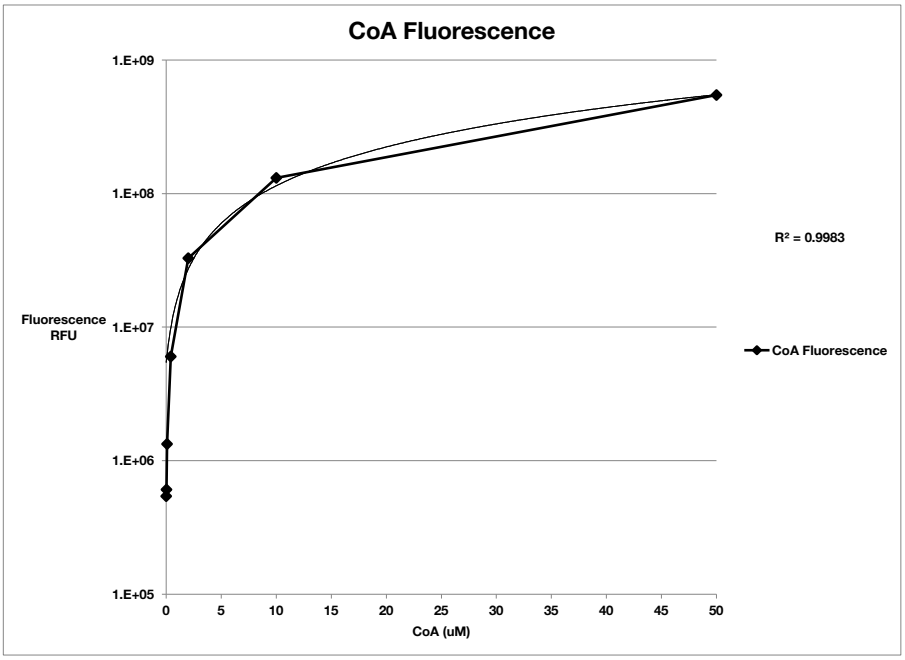



**Figure 3-6: Standard curve of Coenzyme A (CoA) detected using a free thiol fluorescent assay.**
A standard curve is shown using a 5-fold dilution of Coenzyme A from 50 $\mu$M to 170 nm. The Y-axis shows the relative fluorescent units from the fluorescent free thiol detection kit following blank subtraction as well as a negative control, and the X-axis shows the Coenzyme A concentration in micromolar. A linear fit is also shown on top of the observed data as well as sum of the squared residuals on the right hand side.

Linearity of detection was evident from 50 µM to 170 nM CoA . The pH optimum of each of the three PPTases was determined by buffering triplicate reactions in 0.5 unit increments from pH 5.0 to 5.5 using MES, 6.0 to 6.5 using HEPES, and from 7.0 to 8.5 using TRIS at 50 mM each. Reactions were set up as follows: 50 µl total volume with 50 mM buffer, 100 µM CoA, 10 mM MgCl2, 0.2 µM of one of the three PPTases, 4 µM of the BpsA wild-type reporter, and 5 mM ATP with a 10 minutes pre-incubation at 30 ℃ prior to ATP addition. The reaction was allowed to proceed for 20 minutes at 30 ℃ and halted with 250 µl of ice cold 2 M NaCl, 50 mM Tris pH 7.5. Free CoA was removed by passage of the halted reaction through a 3 kD Amicon spin filter at 10,000 × g for 15 minutes at 4 ℃ followed by two washes with 250 µl of 50 mM Tris pH 7.5. The amount of phosphopantetheinate was then determined using the free thiol detection kit according to the manufacturer's directions for each of the triplicate reactions along with three 10-fold dilutions of CoA starting at 25 µM to compare to the standard curve as well as a blank reaction. A pH of 7.5 was chosen for PPTases from clade one and two and 6.5 was chosen for the clade three PPTase based on these empirical results for all subsequent reactions. (Figure 3-7).

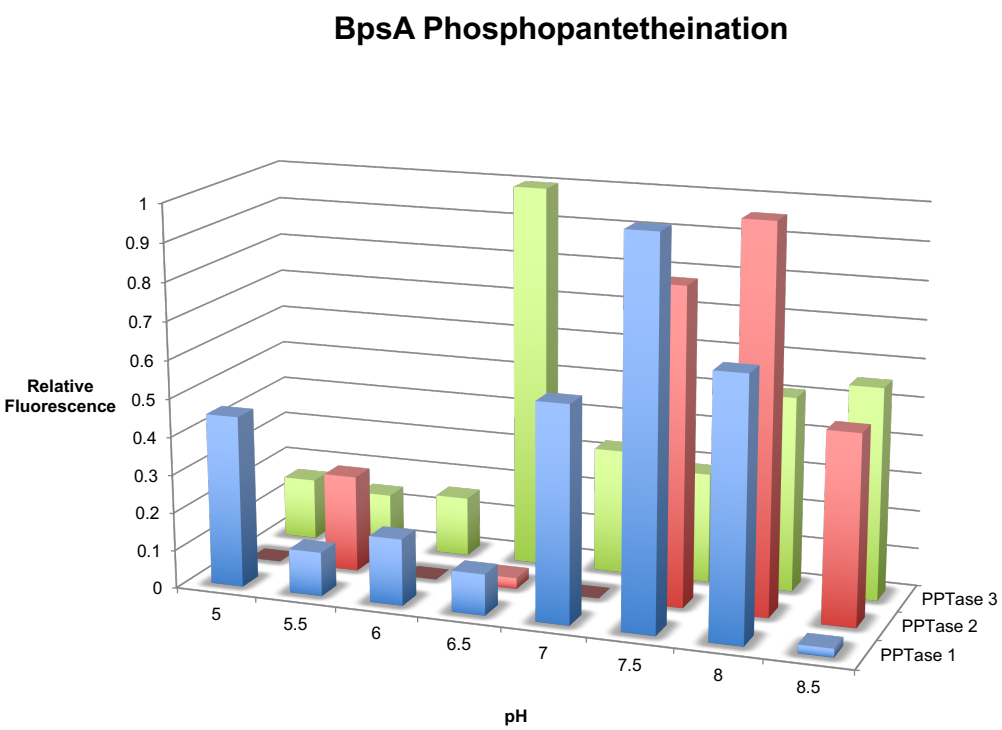



**Figure 3-7: Phosphopantetheination of the BpsA reporter at various pH values**
The graph shows the amount of phosphopantetheinatation of the wild-type BpsA reporter by each of the three *Amphidinium carterae* phosphopantetheinyl transferases. The relative flurorescence produced by the freee thiol detection kit on the Y-axis as a function of pH in the reaction on the X-axis is indicative of the amount of phosphopantetheinate added to the BpsA protein by each phosphopantetheinyl transferase labeled on the Z-axis.

Phosphopantetheination reactions were repeated at optimum pH for each of the three PPTases at 0.2 µM along with the BpsA reporter containing thiolation domains 4 from the triple-KS transcript (3KS4), 1 from the BurA-like transcript (BurA1), and 1 from the ZmaK-like transcript (ZmaK1) at 4 µM, as well as the acyl carrier protein (ACP) at 10 µM in triplicate along with negative controls without CoA added. These reactions were purified using a 3 kD Amicon filter and free thiol was measured using the free thiol detection kit along with three CoA standards and blank controls in the same manner as the pH optimization protocols. Assuming that only one phosphopantetheinate group can be added to each thiolation domain, the amount of free thiol was used to determine the percent of thiolation domains phosphopantetheinated in 20 minutes.

*Results*

Construct generation and domain insertion

Following the generation of the BpsA2.1 vector with restriction sites flanking the phosphopantetheinyl transferase (PPTase) binding site of the thiolation domain, each of the eight dinoflagellate thiolation domain oligonucleotides were successfully inserted and verified by Sanger sequencing in both directions (not shown). Although each of the PPTases from *Amphidinium carterae* were able to interact with the wild type BpsA vector when co-expressed in *E. coli* (Figure 3-2), independent verification of protein production showed very different expression patterns for each of the three PPTases when expressed individually without the BpsA protein (Figure 3-8). In

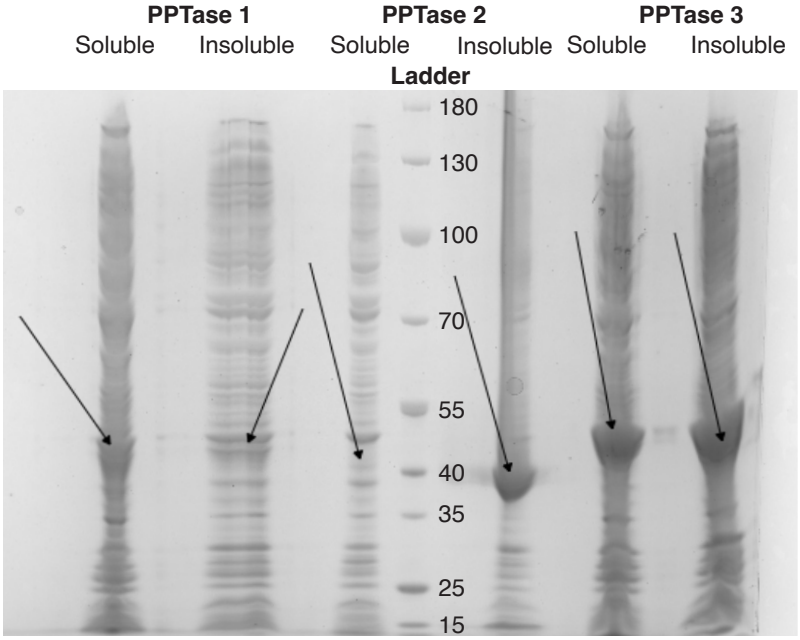



**Figure 3-8: Soluble and insoluble lysates from *E. coli* following induction of phosphopantetheinyl transferase expression.**
An SDS-PAGE gel is shown for three *E. coli* clones conatining the three *Amphidinium carterae* phosphopantetheinyl transferases following induction of protein expression with IPTG. Both the soluble (supernatant following French press isolation) and insoluble (Proteins retrieved from the pellet with 6M urea) fractions are shown with arrows indicating the expected size of each protein based on the molecular weight marker designated as “Ladder” with kiloDaltons indicated.

general, PPTase 3 showed high expression with protein in both the soluble fraction and the insoluble fraction recovered with 6M urea following lysis of the *E. coli* host by French press. PPTase 2, however, was only visible in the insoluble fraction and PPTase 1 had low expression in general.

Each of the constructs produced visible protein following his-tag purification (Figure 3-9), in contrast to the PPTase where PPTase 2 was not present in the

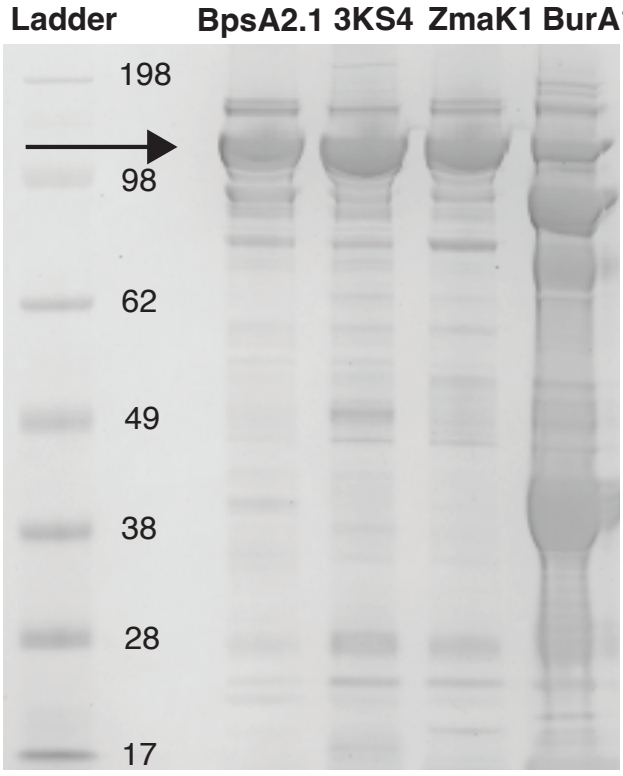



**Figure 3-9: His-tag purified BpsA reporter.** An SDS-PAGE gel is shown for the BpsA2.1 reporter with the standard sequence as well as one each of the triple-KS, ZmaK, and BurA inserts loaded with equivalent total protein. The size marker is shown on the left designated "Ladder" and an arrow shows the expexected reporter size. The BurA1 protein was concentrated prior to imaging and shows several breakdown products.

soluble fraction in appreciable amounts. In order to explain how PPTase 2, which can activate the wild type BpsA reporter (Figure 3-2), can function despite low apparent soluble production in *E. coli*, the soluble lysate from co-expression of PPTase 2 with either the BpsA2.1 vector without a heterologous insert or the ZmaK1 insert were separated by SDS-PAGE (Figure 3-10). The recoverable amount of the PPTase 2

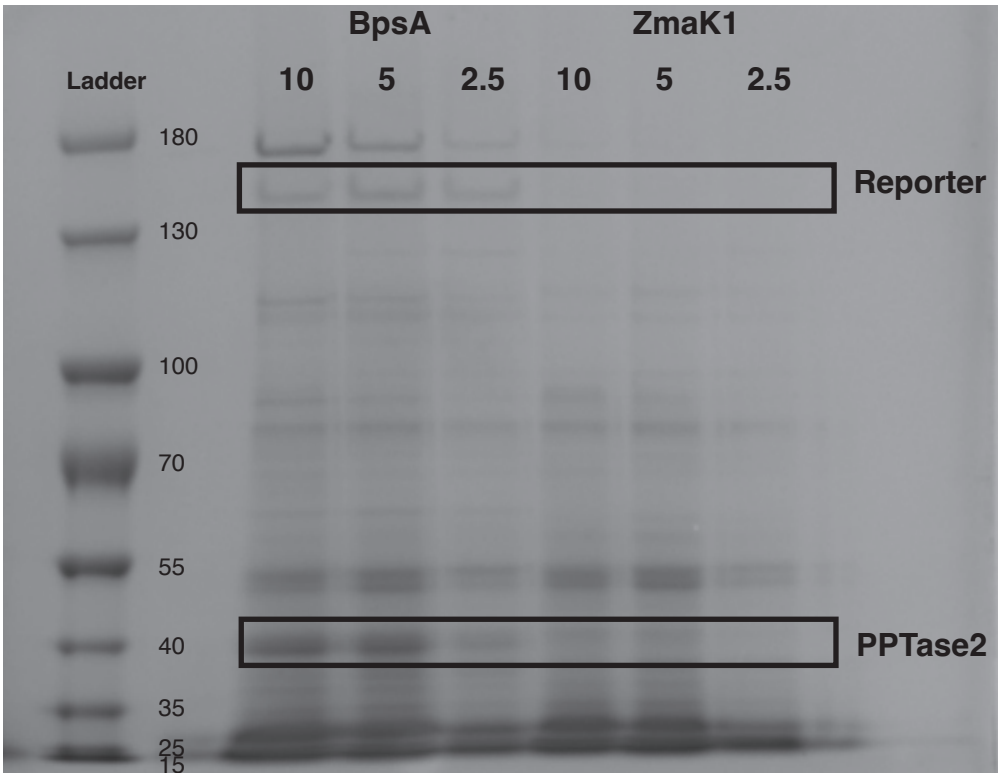



**Figure 3-10: PPTase2 expression with BpsA reporter standard insert and ZmaK insert.**
An SDS-PAGE gel is shown for a co-expression of PPTase 2 with either the standard BpsA2.1 sequence or with the ZmaK1 insert following French press lysis and removable of insoluble material by centrifugation. The expected sizes for the reporter BpsA protein as well as the PPTase protein are highlighted with black boxes according to the expected size shown on the left with the size standard marked as “Ladder”. The load volumes are shown at the top of each well in microliters from equivalent *E. coli* cultures.

protein as well as its substrate BpsA protein were higher in the original vector compared to the ZmaK1 insert containing vector where both the reporter and the PPTase 2 soluble protein were low in abundance.

Indigoidine production in *E. coli*

Following growth on autoinduction plates, co-expression of each of the PPTase activators with one of the BpsA2.1 vectors containing either the modified wild-type sequence or a dinoflagellate sequence insert resulted in similar growth for all colonies but indigoidine production in only some colonies (Figure 3-11).

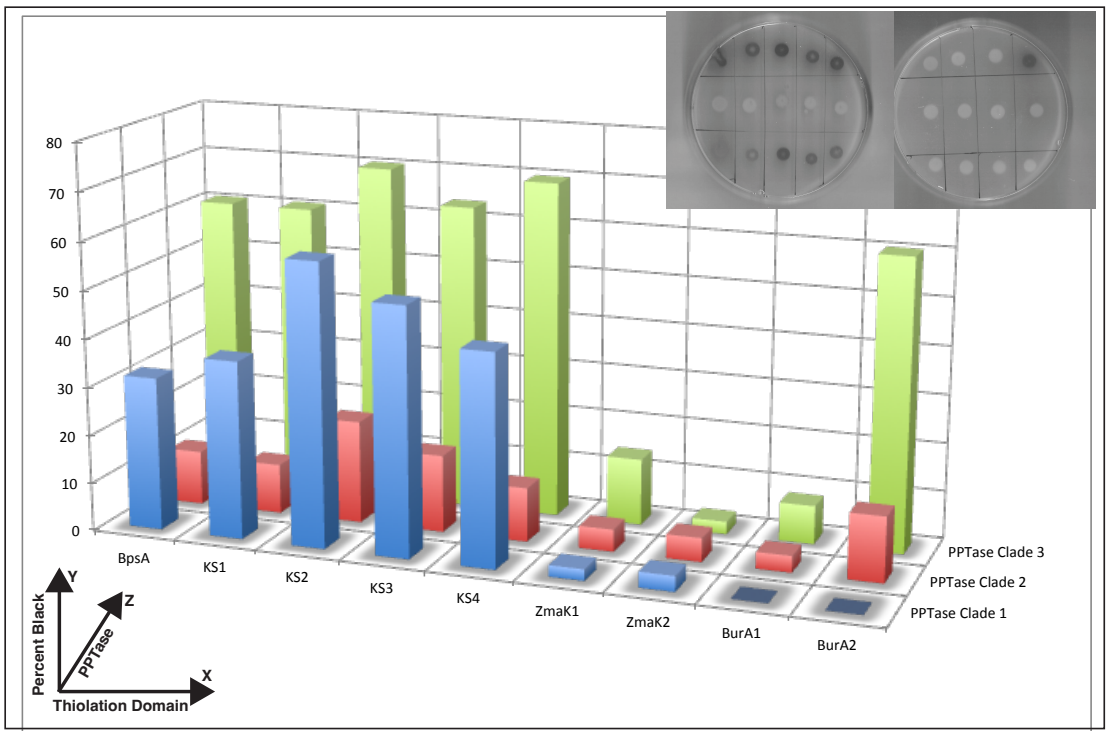



**Figure 3-11: Indigoidine synthesis in *E. coli* from the coexpression of dinoflagellate phosphopantetheinyl transferases and the BpsA gene with a dinoflagellate thiolation domain.**

The graph shows the relative darkness of each colony to a black pixel on the Y-axis resulting from the production of indigoidine in *E. coli* upon the coexpression of the BpsA gene containing a dinoflagellate thiolation domain and a dinoflagellate phosphopantetheinyl transferase (PPTase) on separate vectors. The thiolation domain is indicated on the X-axis including wild type sequence (BpsA), as well as the *Amphidinium carterae* triple KS (KS), ZmaK-like (ZmaK), and BurA-like (BurA) transcript sequences with a numeral indicating which thiolation domain from N to C terminus. The Z-axis indicates which clade of PPTase sequence was coexpressed from Williams *et al.* 2015. The actual plates with induced expression are shown in the upper right in the same orientation as the graph X and Z-axes for reference.

Background subtracted values show higher indigoidine production for the BpsA2.1 vector without inserts as well as with the triple KS inserts but not the ZmaK or BurA inserts with the exception of the combination of BurA2 and PPTase 3. Also, the PPTase 2 activator pairings yielded consistently lower indigoidine production than PPTase 1 or 3 with the exception of the ZmaK2 insert that had low indigoidine production in all cases but was highest with PPTase 2. Other than the ZmaK2 insert, all BpsA pairings with the PPTase 3 activator resulted in higher indigoidine production relative to the PPTase 1 or 2 activators. Indigoidine production was also performed with each insert along with the PcpS gene, a bacterial PPTase from *Pseudomonas aeruginosa* and common control gene for phosphopantetheination, but indigoidine was only produced in appreciable amounts with the reporter without dinoflagellate inserts compared to low or almost no production with the dinoflagellate sequences (Supplementary Figure S3-1).

Thiolation domain phosphopantetheination in vitro

All three PPTases were able to activate indigoidine production in the wild-type BpsA reporter in a dose dependent manner (Figure 3-12). The same was true for

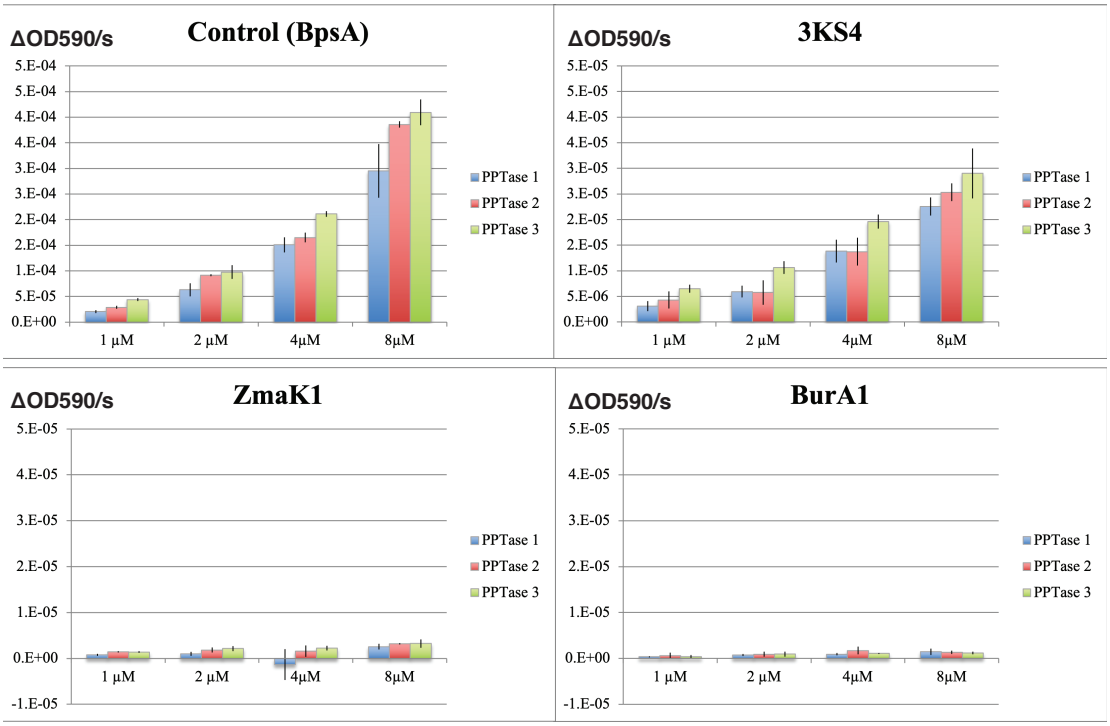



**Figure 3-12: Kinetics of indigoidine production for combinations of thiolation domains and phosphopantetheinyl transfers (PPTases).**
The kinetics of indigoidine production are shown with the change in absorbance at 590 nm on the Y-axis and the concentration of each reporter on the X-axis of each graph titled with the thiolation domain inserted into the reporter. Error bars show the standard deviation for each set of reactions with PPTase clade one in blue, clade two in red, and clade three in green.

thiolation domain four of the triple KS transcript (3KS4) with a similar change in rate versus changes in reporter concentration but with an approximately ten-fold decrease in absorbance at 590 nm. In both cases the clade three PPTase correlated with a slightly higher rate of indigoidine production followed by clade two and then clade one with low error between replicates. Reporters containing thiolation domain one of both the BurA-like (BurA1) and the ZmaK-like (ZmaK1) transcripts did not appear to make appreciable amounts of indigoidine during the time of these experiments for any of the three PPTases.

Phosphopantetheination of the BpsA reporter by all three PPTases was also evident by the fluorescent free thiol detection assay (Figure 3-7). Although the Clade 1 and 2 PPTases had the highest levels of free thiol (phosphopantetheinate) detected at the same pH as the indigoidine synthesizing assays, around pH 7.5, the clade three PPTase had peak fluorescence at pH 6.5. Not surprisingly, the observed phosphopantetheination of the reporter containing the 3KS4 thiolation domain was

much larger with the clade three PPTase than either the clade one or two PPTases when done at the optimum pH (Figure 3-13). This was quantified using a CoA

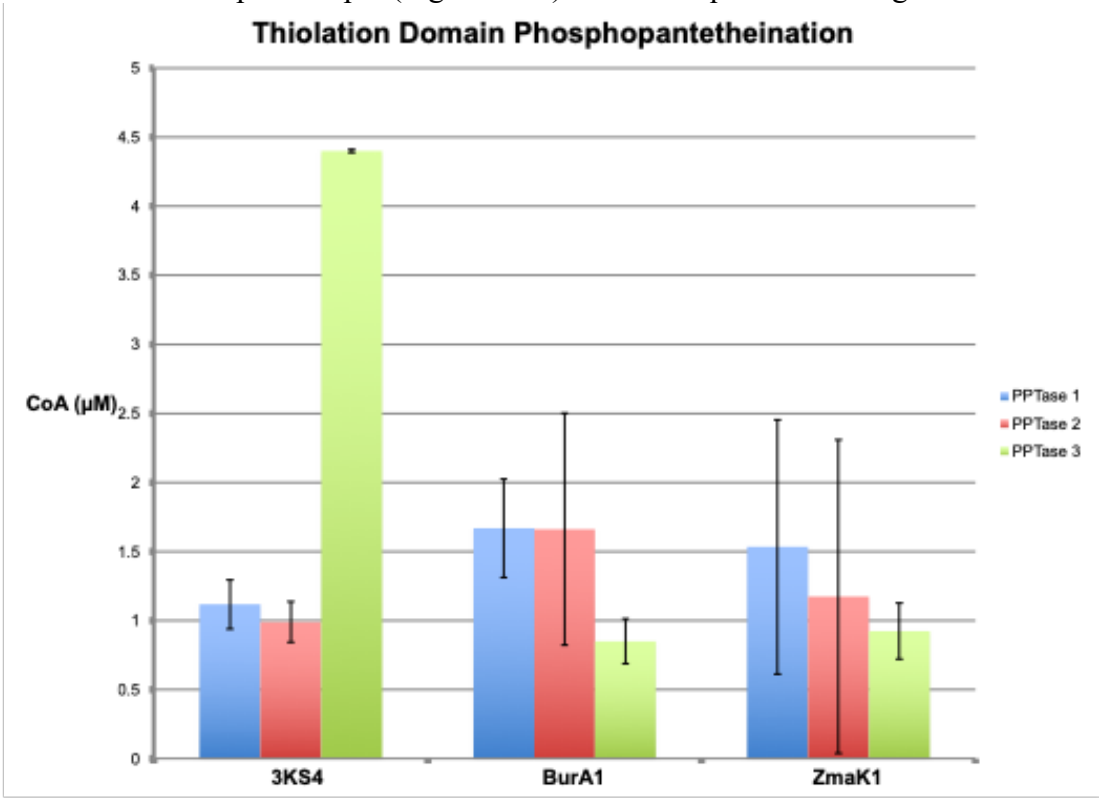



**Figure 3-13: Phosphopantetheination of thiolation domains detected as free thiol**
The graph shows the amount of phosphopantetheinatation of BpsA containing thiolation domain four from the triple KS transcript (3KS4), thiolation dom one from the BurA-like transcript (BurA1), and thiolan domain one from the ZmaK-like transcript (ZmaK1) described on the X-axis. The amount of phosphopantetheination is given as the resultant micromolar amount of CoA added to the 8 $\mu$M of starting BpsA protein. Error bars show the standard deviation of triplicate reactions for the Clade one, two and three PPTases colored blue, red and green, respectively.

standard curve (Figure 3-6) giving a final concentration of CoA attached to the BpsA protein in µM. Relative to the 8 µM starting material of BpsA gives a phosphopantetheination level for the 3KS4 domain of approximately 50% for the clade three PPTase and approximately 12.5% for the clade one and two PPTases. Despite a lack of evident phosphopantetheination in the indigoidine production assays for the BurA1 and ZmaK1 thiolation domains, phosphopantetheination was evident in the free thiol detection assay for these domains for all three PPTases but with a larger error. Also, the Clade 3 PPTase gave the lowest yield for BurA1 and ZmaK1 with approximately 10% of starting material phosphopantetheinated while the Clade 1 and 2 PPTases yielded approximately 18% phosphopantetheination.

Phosphopantetheination was also evident using the free thiol detection method for the acyl carrier protein (ACP), the thiolation domain responsible for lipid synthesis (Figure 3-14). Again, all three PPTases were able to phosphopantetheinate

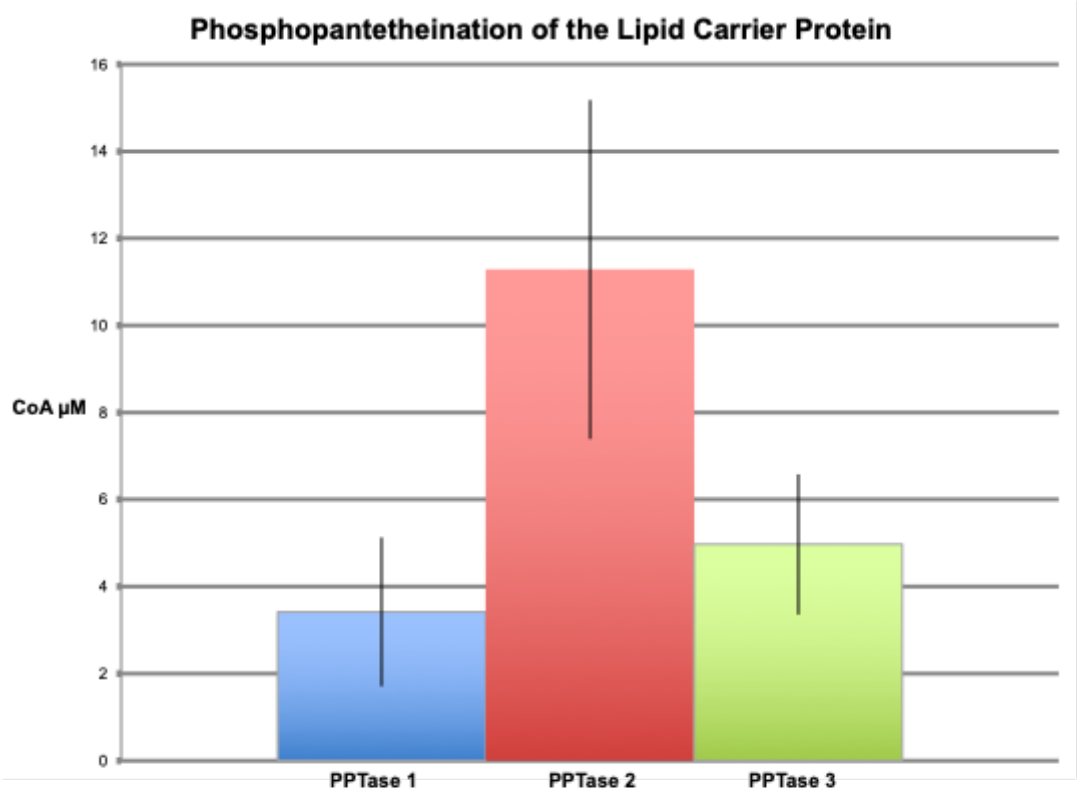



**Figure 3-14: Phosphopantetheination of acyl carrier protein detected as free thiol**
The graph shows the amount of phosphopantetheinatation of the acyl acrrier protein (ACP), the thiolation domain for lipid synthesis. The amount of phosphopantetheination is given as the resultant micromolar amount of CoA added to the 15 $\mu$M of starting ACP protein. Error bars show the standard deviation of triplicate reactions for the Clade one, two and three PPTases colored blue, red and green, respectively.

this thiolation domain with the clade two PPTase yielding the highest level with approximately 75% of the 15 µM starting material phosphopantetheinated followed by clade three at 30% and clade one at 23%. This free thiol detection method cannot be compared to the indigoidine synthesis assay for the ACP since the ACP containing BpsA gene was not expressable in *E. coli*.

## *Discussion*

### The production and use of recombinant dinoflagellate proteins

This work is the first example of heterologously expressed dinoflagellate proteins used in an in vitro assay. It is also the first example of a catalytically active dinoflagellate protein produced by an in vitro synthesis method. This is quite exciting given the general difficulties in heterologous protein expression (Rosano &

Ceccarelli, 2014; Sørensen & Mortensen, 2005), but also means that codon optimization is not necessarily required (Angov et al., 2008), allowing for the use of native dinoflagellate transcripts. *E. coli* was used successfully in this study for protein expression, but mammalian cells have also been used (Ma, Shi, & Lin, 2020) indicating that eukaryotic cells can also be utilized for proteins that require cleavage, chemical modifications, or to test cellular localization predictions. The use of in vitro and in-vivo methods will hopefully increase our understanding of dinoflagellate biology where common techniques such as promoter modification and targeted genetics have been hampered by a largely post-transcriptional control of gene regulation and very high gene copy number (Lidie et al., 2005; Morse et al., 1989; Roy et al., 2018; Bachvaroff & Place, 2008).

Gene knockdowns and knockouts are the most common method for modifying protein expression in dinoflagellates (Diao, Song, Zhang, Chen, & Zhang, 2018; Yan, Wu, Kwok, & Wong, 2020) and are likely the methods moving forward for understanding natural product synthesis in dinoflagellates, despite the difficulties presented by high copy number. This study exploits the stable chemical modification of one actor in natural product synthesis by another. This allowed us to test the interactions of two proteins indirectly by measuring indigoidine synthesis or directly by measuring the increase in free thiol groups following the attachment of the phosphopantetheinate group to the thiolation domain. In the downstream reactions the biosynthetic enzymes interact with the chemical being synthesized either directly or at the site of attachment to the phosphopantetheinate group making specific interactions much more difficult to discern. With dozens of ketosynthases to choose from, trying to figure out which one attaches which exact acetate in a molecule like amphidinol with sixty-five carbons (Houdai et al., 2001) is certainly daunting. Thus, indirect methods may be more useful such as targeting acyl transferase and thioesterase domains that, as noted by Van Wagoner *et al.* (Van Wagoner et al., 2014), are more likely to have a recognizable impact on the final structure of dinoflagellate toxins and potential inhibitors exist for these enzymes (Lupien et al., 2019; Naik et al., 2014; Valastyan et al., 2020). They are also generally low in copy number similar to the PPTases used in this study with an average of twelve thioesterase domains in dinoflagellate transcriptomes (Table 1-2) (Williams et al., 2021). In addition to being the target of PPTases, the thiolation domains likely involved in natural product synthesis have been shown to have six to seven tetratricopeptide repeats (Clairfeuille et al., 2015), unlike the acyl carrier protein (ACP) for lipid synthesis, that may serve to scaffold biosynthetic complexes. This could allow for the enrichment of catalytic complexes for natural product synthesis from protein extracts using antibody-based methods. Essentially, a bit of whittling down is necessary before further biochemical validation of the roles of dinoflagellate natural product genes is feasible.

The differences between the indigoidine production and the free thiol detection assays show a disconnect between the ability of the BurA and ZmaK insert reporters to produce indigoidine and their ability to be phosphopantetheinated (Figs. 3-11, 3-12, 3-13). The lack of indigoidine production is likely due to steric inhibition from the thiolation domain that must be positioned to interact with all other domains for successful biosynthesis (Figure 3-2). This may help explain why there is a reduction

in indigoidine synthesis with the 3KS4 thiolation domain versus the original BpsA domain (Figure 3) but with a similar change in rate with increasing substrate concentration. This makes the indigoidine production assay useful for providing yes/no type answers as has been suggested before (Owen et al., 2011) but requiring confirmation of negative results. The free thiol assay on the other hand is a direct measurement and likely at least semi-quantitative. To be adapted to other types of natural product synthesis interactions, a detectable chemical change is required, which is unfortunately hard to come by. Modified analogs or radioisotopes have also been used in the past for PPTase (Bunkoczi et al., 2007; Cai et al., 2005; Sonnenschein et al., 2016), but may be difficult to extend to the downstream biochemistry of natural product synthesis given that acetate, the dominant substrate incorporated into dinoflagellate toxins, is used in so many other biological processes. Development of an in vitro assay would alleviate this but may prove tedious given the large number of ketosynthases in dinoflagellate genomes.

### Phosphopantetheinyl transferase/thiolation domain specificity and evolution

There were some obvious differences observed between the triple-KS inserts and the ZmaK or BurA inserts in terms of the indigoidine produced (Figs. 3-11 and 3-12). The triple-KS inserts had consistently high indigoidine production, especially with the ubiquitous Clade 3 PPTase, and the triple-KS transcript can also be found in the more basal syndinian dinoflagellate *Hematodinium sp.* (Gornik et al., 2015), a parasite of crustaceans. The BurA-like and ZmaK-like genes on the other hand are not found in any syndinian transcriptomes to date and are very similar in sequence to bacterial genes making horizontal transfer a likely origin. The results presented here may indicate that, at least for the *Amphidinium carterae* PPTases, the ability to activate the BurA and ZmaK inserts is sub-optimal. This is also evident in thiolation domain sequence clusters based on the observation that many of the ZmaK sequences lie outside the cluster of natural product associated domains with the more basal sequence the furthest away, indicating that convergent evolution may be an active force (Figure 1-9). Thus, PPTases from more distal species of dinoflagellates may be better at phosphopantetheinating the BurA and ZmaK thiolation domains than the *A. carterae* based PPTases used here.

The results of the free thiol assay show that each of the three PPTases can phosphopantetheinate all of the thiolation domains used in this study, including the acyl carrier protein at near equivalent amounts (Figs. 3-13 and 3-14). This is intrigui9n given that the Clade 2 PPTase exists in the genome without having an apparent stop codon and without evidence of expression as an intact protein (Figure 2-5). This is especially surprising considering that the clade two PPTase was able to phosphopantetheinate the ACP, required for the vital process of lipid synthesis, to a greater extent than the other two PPTases (Figure 3-14). The second surprising interpretation based on the apparent lack of PPTase specificity is that lipid and natural product synthesis are not segregated based on PPTase targets as is usually the case in bacteria and fungi (Beld et al., 2014; Gerc et al., 2014). The most basally branching dinoflagellates are the syndinian clades (Bachvaroff et al., 2014) that are not

photosynthetic but instead parasitize other eukaryotes. Although not well studied, there is some genetic information including genomic data for *Ameobophyra ex. Karlodinium veneficum*, a parasite of the dinoflagellate *Karlodinium veneficum* that does not have any lipid synthesis machinery and presumable gets its lipids from the host (Bachvaroff, 2019). This includes the absence of a PPTase, an enzyme generally assumed to be present in all life since lipid synthesis is usually required for biological life to exist (Beld et al., 2014). Another species of syndinian dinoflagellate is *Hematodinium sp.*, a crustacean parasite that also does not have an identifiable PPTase but does possess the triple KS transcript from which one of the thiolation domains in this study is derived (Gornik et al., 2015). Thus, it is unlikely that dinoflagellates have a native PPTase and the PPTases that are evident may have been acquired through horizontal gene transfer, possibly from the chloroplast as has happened in the other cases (Dorrell & Howe, 2015; Ishida & Green, 2002; Yamada et al., 2019). This may explain the lack of binding site specificity observed with these PPTases as well as other protists (Sonnenschein et al., 2016). Whether specificity was present when acquired and deteriorated over time or if the PPTase acquired during endosymbiosis already lacked specificity is unclear. The benefit of this lack of specificity is that natural product genes can be acquired over time and are more likely to be utilized if there is an existing PPTase that can activate the biosynthetic pathway. This may help to explain the likely horizontal gene transfer of the BurA-like and ZmaK-like genes that are not evident in any of the basal syndinian lineages but are common in most core dinoflagellate lineages. Also, while there are three PPTases in *Amphidinium carterae*, a basal core dinoflagellate, there are many derived species that have lost PPTases, including *Protoceratium reticulatum* that only has the clade three PPTase (Williams, Bachvaroff, & Place, 2020). Thus, natural product synthesis in dinoflagellates may be different bacteria and fungi because the biosynthetic capabilities have been acquired by horizontal gene transfer.

## *Conclusion*

The demonstration of using purified protein in in vitro assays is an important advancement in the search to understand dinoflagellate biology. They are such a strange and diverse group of organisms that comparisons to other species, especially many common models, can be misleading. Even in this study, the interpretation of the in vitro assays alone might lead one to assume that the clade two PPTase is most likely responsible for activating lipid synthesis by phosphopantetheinating the acyl carrier protein. Looking at the protein expression gives the very unexpected result that this protein is seemingly immediately broken down making a biological role for this enzyme entirely unlikely apart from a possible regulatory agent. Indeed, this study utilized enzymes from *Amphidinium carterae*, but proteins from other species may give very different results and may have unique functionalities. Taking as holistic an approach as possible is important when dealing with such a dynamic group of organisms, and biochemical validation will certainly be an important tool in the

future. This study also reveals the importance of taking the evolutionary history of dinoflagellates into account when interpreting data. Although the bulk of study is on the core dinoflagellates that are dominated by photosynthetic species, the common ancestor is likely heterotrophic and parasitic. This means that any plastid associated process is likely to have a very complicated evolutionary history and keeping an open mind is essential.

# Overall Conclusions and Future Work

*Summary of results*

The overarching goal of these studies was to identify and characterize genes whose function could distinguish lipid synthesize from toxin synthesis in dinoflagellates. A hidden Markov model (HMM) was successfully developed that could identify a variety of candidate genes potentially involved in lipid and toxin synthesis. For some domains the genes likely involved in lipid synthesis could readily be identified such as thiolation and ketosynthase domains whereas in ketoreductase and acyl transferase domains the gene count was drastically expanded leaving the concept of functional segregation in question. The gene counts in general are higher than expected and the domains themselves do not follow the mathematical canon where domains that act downstream are as abundant or lower in abundance than the upstream elements. It would be like having more apple juice factories than you have apple farms, leaving open the possibility that some elements in dinoflagellate natural product synthesis may act iteratively or in multiple processes. The structural elements that would in theory scaffold these synthetic elements together were also discovered on thiolation and acyl transferase domains. These particular domains changed in relative abundance during core dinoflagellate evolution with an increase in acyl transferase domains and a decrease in thiolation domains indicating a change in strategy for how synthetic complexes are organized. The thiolation domains were chosen for future study because of their apparent sequence segregation with respect to lipid and natural product synthesis, the presence of scaffolding domains, and because they are acted on very early in synthesis by phosphopantetheinyl transferases (PPTases).

The phosphopantetheinyl transferases themselves were shown to be in low copy number in dinoflagellates with up to three or as few as one copy present. They were also found to be distinct from PPTases in bacteria and fungi making immediate functional classification difficult. A lack of targeting motifs further complicates matters since nearly all of the domains that they act upon have clear chloroplast targeting sequences. The clade one and three proteins in *Amphidinium carterae* were expressed in an alternating pattern over both a day night cycle as well as over a growth curve while the clade two protein was strangely never observed in its whole form. Instead it was detected as a breakdown product common to all three PPTases indicating a possible mechanism for inactivation. The clade two PPTase was found to lack a stop codon providing an explanation for its seeming immediate degradation and the 3′ sequence was similar to the clade one sequence hinting at a likely duplication and subsequent loss of function. The acyl carrier protein on the other hand was quite standard in both expression pattern and sequence motifs.

The clade one and three PPTases were successfully expressed in *E. coli* and purified as active protein. The clade two PPTase was not able to be expressed as soluble protein in *E. coli* and instead was synthesized in vitro and purified. All three PPTases, however, were able to be expressed in *E. coli* along with the indigoidine synthesizing gene BpsA from *Streptomyces lavendulae*. This gene requires activation by a PPTase for indigoidine production and the *E. coli* host doesn't not possess a PPTase capable of activating the BpsA protein. All three *Amphidinium carterae* PPTases were able to activate the BpsA protein to varying degrees implying the functional ability to activate thiolation domains involved in the synthesis of other natural products like toxins. The BpsA gene was successfully modified to allow the incorporation of dinoflagellate thiolation domain sequences by the addition of unique restriction sites. Eight thiolation domains from three dinoflagellate transcripts as well as the acyl carrier protein were all incorporated into the BpsA gene. All were expressed in *E. coli* and purified and purified as active protein with the exception of the acyl carrier protein, likely due to host toxicity. This particular protein was synthesized in vitro and purified. The purified BpsA proteins as well as purified PPTases were allowed to react in vitro with indigoidine production using all three PPTases evident from the thiolation domains of one of the dinoflagellate transcripts but not the other two. This strange result was double checked by fluorescent detection of the free thiol resulting from the activity of the PPTase on the thiolation domains. This method was able to demonstrate that all of the thiolation domains, including the synthetic acyl carrier protein, were all able to be phosphopantetheinated by all of the PPTases and that the lack of indigoidine synthesis was likely the result of the dinoflagellate sequence disrupting the activity of the BpsA gene.

*Successful development of in vitro assays to test the interactions of dinoflagellate proteins*

In order to test sequence-based predictions of protein products it is necessary to test the protein function in a controlled environment. These experiments exploited the direct interactions of two proteins involved in dinoflagellate natural product synthesis, the thiolation domain and the phosphopantetheinyl transferase (PPTase). The most important aspect of this interaction is that the thiolation domain is transformed from the apo form to the holo form by the PPTase. This means that the functionality of the thiolation domain is unlocked by the PPTase and that without this conversion the thiolation domain is entirely inactive. In effect the PPTase is a switch that turns the thiolation domain on. Thus, there are two ways of approaching this particular interaction: direct assessment of thiolation domain modification by the PPTase and functional activation by the PPTase

In these experiments direct assessment of thiolation domain modification was performed using a fluorescent assay that detects free thiol groups. This is convenient for this particular modification because the PPTase adds a moiety with a single free thiol group making quantification fairly straightforward. Another version of this technique is to use radioisotopically labeled CoA and measure the subsequent

increase in radioactivity of the thiolation domain following phosphopantetheinate transfer. The use of radioisotope is extremely sensitive and quantitative when done right. $^{32}$P labeled CoA is commonly used with the consideration that either both phosphates are labeled or the phosphate distal to the nucleotide to ensure that radioactivity is conferred to the thiolation domain. $^{35}$S is another possibility although not as easily detectable as $^{32}$P. The advantage is that there is a single sulfur to trace. The use of radioisotope lends itself to in vitro assays but not pulse chase studies since CoA is used in so many reactions potentially muddying subsequent results. Another method to directly measure phosphopantetheination is the "P-eject" method that is a tandem mass spectrometry based method. The phosphopantetheinated thiolation domain is infused into the mass spectrometer in the case of in vitro reacted material or a biological lysate is passed to the mass spectrometer following purification and liquid chromatography for in-situ reactive material. Parent molecules with masses equal to the phosphopantetheinated thiolation domain are passed through a collision gas where the phosphopantetheinate group is preferentially removed and then detected in the second spectrometer. This is a definitive way of establishing that the protein has been phosphopantetheinated but potentially difficult to make quantitative, requiring a significant amount of time for method development. In general, these methods could be used in future experiments whenever a substrate is modified by an enzyme that adds a distinct chemical group. Some examples are protein ubiquitination and phosphorylation, carbohydrate acylation, or possibly nucleic acid cleavage. There are considerations on whether to use radioisotope, antibody, mass spectrometry, or another method for detection but all of these examples can employ heterologously expressed protein in vitro using the methods developed here.

While direct measurements can work with the pairing of PPTases and thiolation domains, many other genes that participate in natural product synthesis have a more nuanced role and don't directly modify their substrate. Instead, their role is dictated by the substrates they interact with and at what point in the synthetic process they are recruited. The point of recruitment can be quite difficult to ascertain in dinoflagellates since, as stated before, their gene order does not reflect the synthetic order and the gene copy number is very large. This makes knockouts a much more relevant approach followed by a survey of product intermediates to see where synthesis was disrupted. The experiments here focused on substrate specificity by performing domain replacement in a synthetic system with a known product. This is especially useful when dealing with PPTase specificity because the final natural product isn't modified by domain replacement. The result of the domain swap is instead whether the product is made and how much. This was successful for some dinoflagellate domains although incorporating others appeared to disrupt the ability of the reporter system to produce indigoidine demonstrating the greatest limitation of this method. The advantage on the other hand is the ease of domain replacement. Swapping nucleic acid sequences is trivial compared to the expression and purification of active protein. This allows for many domains to be tested in a single round of experimentation. A primary target for future experimentation would be the adenylation domains in dinoflagellates. While a few of them are associated with condensation domains, the vast majority have no cognate condensation domain to perform traditional peptide bond formation despite their commonly predicted

substrates of cysteine and phenylalanine. Using domain swapping to determine the possible substrates of these domains could answer a lot of questions about what natural products dinoflagellates make and how they make them.

*Dinoflagellate natural product and lipid synthesis domains*

These experiments failed to use PPTases to differentiate classic lipid synthesis from natural product synthesis because in some ways they are not differentiated, at least not in the canonical sense. Given the fact that the syndinian dinoflagellates did not make lipids and instead assimilated them from their host it appears that dinoflagellates acquired both lipid and natural product synthesis during chloroplast acquisition. This is drastically different from bacteria and fungi that had mechanisms for regulating lipid synthesis prior to the evolution of natural product synthesis. Thus, it is not too surprising that elements of lipid and natural product synthesis have been jumbled up in dinoflagellates in a novel way. All of the pertinent elements for natural product and lipid synthesis appear to be translocated to the chloroplast based on targeting sequences, except for the PPTases, which is quite unusual. It could be that all thiolation domains are phosphopantetheinated before they even reach the site of synthesis such as in the Golgi or even in the endoplasmic reticulum. This would certainly be in line with the lack of substrate specificity apparent in dinoflagellate PPTases if phosphopantetheination is coordinated with all possible processes rather than segregated to particular avenues of biosynthesis. Certainly, microscopy using specific antibodies can help determine where the PPTases are expressed to answer questions about their role. Guessing which antibodies to use may prove tricky since the expression of each PPTase is highly variable.

So, if lipid and natural product synthesis are not separated by PPTase specificity as in other organisms then how are these processes regulated? One clue is the acyl carrier protein. Its expression pattern changes with the stage of growth and may be the primary limiter for the rate of lipid synthesis. Also, the acyl carrier protein does not have any of the scaffolding domains that the other thiolation domains and the acyl transferase domains possess. Thus, gene recruitment in dinoflagellate natural product synthesis may play a huge role in what the final molecule is, helping to explain why there are so many different types of dinoflagellate natural products. The sequence differences between the acyl carrier protein and other thiolation domains that were observed here may influence which genes are recruited to the complex and not PPTase binding as was original theorized. Some evidence for this includes the ketosynthase domain, the first enzyme to interact with the carrier domain following phosphopantetheination. This is the only other domain where the gene copies expected to participate in lipid synthesis are very well conserved and in low copy number. The acyl carrier protein and the ketosynthase domain may help recruit domains that are specific for lipid synthesis ensuring complete saturation of acyl chains. This could be tested by expressing the ACP and some candidate ketosynthase domains in vitro using the same techniques presented here along with radiolabeled malonyl CoA and look for incorporation of isotope. Differential incorporation would

be an unprecedented example of coordination between a carrier protein and a ketosynthase domain. Otherwise, the strict conservation observed in the ketosynthase domains, but not other domains would be difficult to explain. Performing a knockout of the dehydratase domains may also prove fruitful since they are also in very low copy number and there are very few double bonds observed in dinoflagellate toxins leaving polyunsaturated fat and saturated fat synthesis as the most likely role for these enzymes.

Separating polyunsaturated fat synthesis from other natural products is tied to the mystery why dinoflagellates have both multidomain and single domain transcripts. While polyunsaturated fats can theoretically be synthesized by any number of domains, the presence of the triple-KS transcript in the syndinian dinoflagellate *Hematodinium sp.* is striking and indicates that this transcript may make polyunsaturated fats since *Hematodinium* does not make any known toxins. On the other hand, the triple-KS could just as easily synthesize major portions of the backbone structure of most dinoflagellate toxins by iteratively using each module, indicating a separate rationale for its conservation. The questions remaining for the triple-KS and other multi-domain transcripts are whether they process natural product synthesis iteratively and if so how much. One way to get at this would be to use thioesterase inhibitors to prevent cleavage of the natural product and examine what molecules are still attached to the thiolation domains. This could be done heterologously assuming the appropriate substrates are available or in-situ. The complex could then be pulled down using specific antibodies. One clear advantage is that multiple different antibodies could be used for multidomain transcripts preferentially pulling down these proteins versus single domains. Using radiolabeled substrate would not be advisable in this case since acetyl-CoA is used in so many processes and may be hard to trace. Rather, excision of the natural product from the thiolation domain and mass spectrometric analysis is preferable to characterize the products of the multidomain transcripts. This would help to answer questions about whether these multidomain transcripts serve to make polyunsaturated fats, toxins, or both.

*Gene expansion and retention in core dinoflagellates*

Altered chromosome structure, a high retention rate of horizontally transferred genes, post-transcriptional control of gene expression, and retention of a chloroplast are all synapomorphies of core dinoflagellates. Why these traits all co-occur is not entirely clear but the results of these studies may surprisingly shed some light. We know that the appearance of alternate proteins involved in chromatin structure appear in late branch syndinian dinoflagellates. In core dinoflagellates we see a big increase in gene copy number, possibly from the increased likelihood of recombination that followed a change in chromatin structure. More importantly is a shift away from transcriptionally controlled gene expression like most other organisms and the emergence of a system where the default is for translation to be proceed unless translation is inhibited through regulatory elements, mostly circadian. This may be what allowed for the retention of horizontally transferred genes and ultimately

endosymbiosis of photosynthetic prey because any novel transcripts that do not have regulatory elements will be translated. This means that any potential benefit will be realized from the start with selection pressure acting to dial back and regulate expression rather than enable it. This is how core dinoflagellates were suddenly able to synthesize lipids, make polyketide based natural products, conduct photosynthesis, and many other metabolic processes that their chloroplast could do when it was a free living haptophyte. Once the genes for these processes make their way to the nucleus they can expand like other dinoflagellate genes, which is why we see several redundant PPTases, a huge repertoire of polyketide synthesis genes that allow for toxin and polyunsaturated fat synthesis, and an acyl carrier protein more similar to bacteria than closely related eukaryotes. Thus, we see a sudden gain followed by many separate losses of PPTases and likely other genes as well (Figure C-1).

So then why is PPTase two expressed and immediately degraded? Why are so many other proteins expressed and degraded on a circadian cycle? Woody Hastings postulated a theory that it was a way of recycling nitrogen. Although this ideas was never fully flushed out, it makes sense when you consider that the acquisition of a chloroplast was likely the most dramatic, beneficial, and dangerous event in dinoflagellate evolution. Splitting water and releasing electrons can be incredibly harmful effectively burning the cell to death if not controlled properly. The final electron acceptor is NADP, one of the three major nitrogenous compounds in the cell along with chlorophyll and amino acids. If redox stress gets too high, amino acids are a ready source of nitrogen and it pays to keep them around. Most if not all dinoflagellates are heterotrophic at some level meaning that they can ingest nitrogen. Food isn't always available, though, and in lean times the constant synthesis and degradation of protein may be a way of storing nitrogen in the cell in a manner that is useful but also accessible. The dinoflagellate toxins themselves have been shown to be associated with sensing redox stress in *Karenia brevis* and this may be a common use in most other core dinoflagellates. Toxins may be both a sink for electrons where free NADP can be recycled as well as a means of sensing when oxidative stress is too high. Except in *Karlodinium veneficum*. In *K. veneficum* the toxin is released for prey capture because if you are extra effective at catching prey you don't have to worry so much about available nitrogen. We can envision a conceptual model where the biology of core dinoflagellates that allowed for chloroplast acquisition and horizontal gene transfer made toxin and lipid synthesis possible. It also explains the diversification and expansion of these processes in ways that don't occur in most other organisms. Many of the methods employed in these studies can be used and expanded upon as discussed to decipher the pathways of toxin synthesis in core dinoflagellates, but each result must be taken lightly because we can expect unique changes in most if not every species examined, just as we've seen with PPTase evolution.

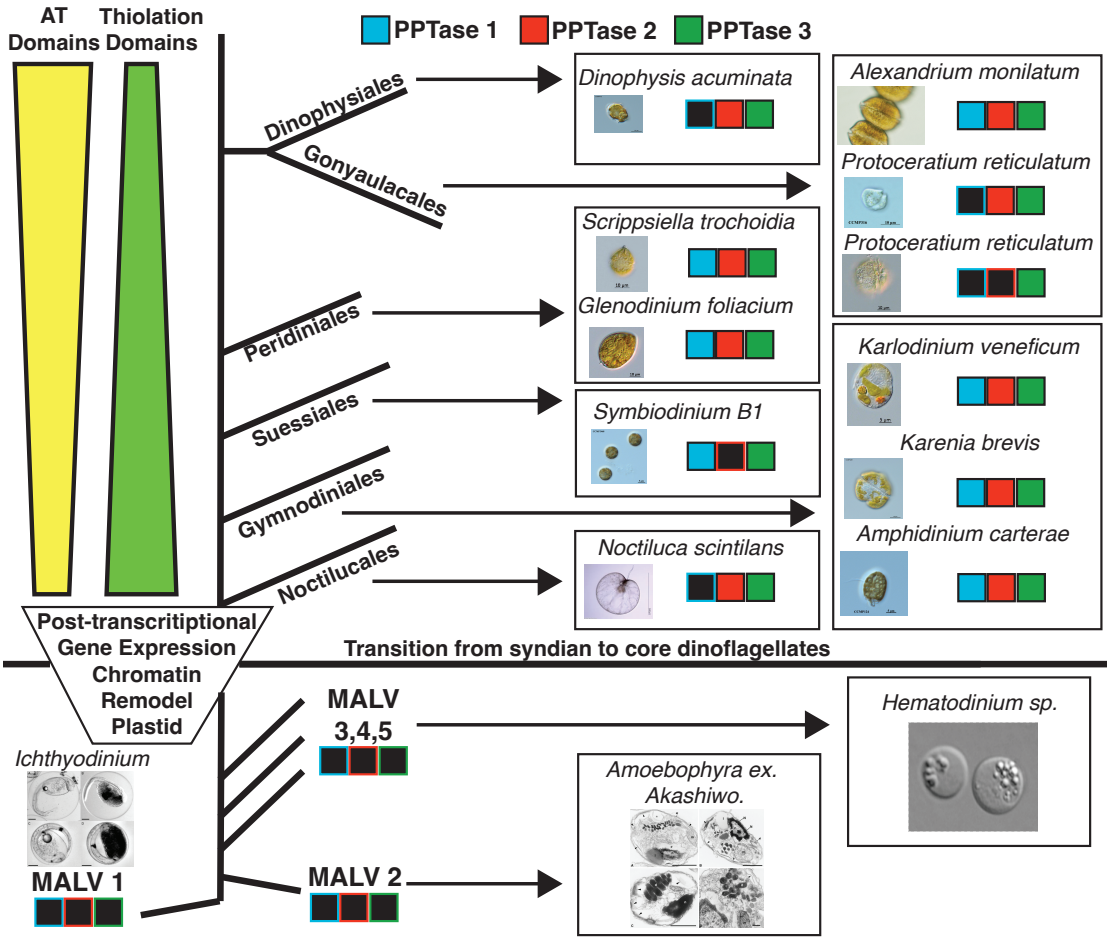



**Figure C-1: Natural product domain evolution in dinoflagellate**
Shown is a cartoon of dinoflagellate evolutionary groups with a horizontal black line separating the syndinian dinoflagellates on the bottom from the core dinoflagellates on the top. Example species of Order level taxonomies are shown in black boxes next to a corresponding arrow along with boxes showing the presence or absence of each of the three PPTase clades. A colored box with a black border shows the presence of a PPTase clade while a black box with a colored border indicates the absence according to the legend at the top. The cartoon to the upper left represents the increase in acyl transferase (AT) domains and the reduction in thiolation domains during core dinoflagellate evolution.

# Appendices

A1: Spliced leader characterization

**Spliced leader containing sequence enrichment and 5′ cap isolation**

Twelve 4-liter samples of an exponentially growing *A. carterae* culture were taken twice weekly, pelleted at 1000 x g for 10 min and frozen at -80° C. Each pellet was used for RNA extraction using the RNAzol RT reagent (Molecular Research Center, Cincinnati, OH) according to the manufacturer's instructions. Briefly, pellets were homogenized in the RNA extraction reagent, the nucleic acids were allowed to deproteinate at room temperature for 15 min, and cellular debris was removed by centrifugation. Large RNAs (approximately >200 bases) were precipitated with 0.4 volumes of 75 % ethanol and pelleted by centrifugation. The remaining small RNA fraction was precipitated from the supernatant with 0.8 volumes of isopropanol and pelleted by centrifugation. The pellets were suspended in water treated with diethyl pyrocarbonate (DEPC) and quantified. The isolated RNA from each size fraction was pooled and re-precipitated with an equal volume of isopropanol at room temperature for 15 min, pelleted at 12,000 x g for 20 min, washed with 75 % isopropanol, and re-pelleted at 12,000 x g for 5 min. The final pellets were suspended in 200 µl of DEPC water, yielding approximately 200 µg of RNA >200 bases and 5.2 µg of RNA <200 bases.

For the U4 and SL pulldowns, 1 µg of the small RNA fraction was diluted to 44 µl and combined with 5 µl of 5 M NaCl and 1 µl 1 M $MgCl_2$, each. The RNAs from the >200 base fraction were similarly diluted and also used for spliced leader enrichment. 50 µl of formamide was added to each 50-µl sample and combined with 100 µl 2X hybridization buffer (8X SSC, 1mM EDTA, 20 % dextran sulfate). 400 µl of streptavidin coated beads were washed with washing buffer and bound to the biotinylated primers (Table 3) for 1 h at room temperature on a rotisserie. The RNA samples in hybridization buffer were combined with 100 µl of the bead bound oligonucleotide and hybridized for 18 h at 40° C. The beads were washed five times with 500 µl washing buffer. The SL pulldown from the large RNA fraction was suspended in 100 µl RNAse A buffer (10 mM Tris-HCl, pH 7.6, 1 M NaCl). The SL pulldown from the small RNA fraction was suspended in 100 µl of RNAse T2 buffer (10 mM ammonium acetate, pH 4.5, 10 mM EDTA), and the U4 sample was suspended in 100 µl decapping buffer. 1 µl of 10 mg/ml RNAse A (Thermo, Waltham, MA) was added to the large RNA SL pulldown and single stranded RNA was degraded at 37 °C for 1 h while 1 µl recombinant RNAse T2 (Mo Bi Tech, Goettingen, Germany) was added to the small RNA SL pulldown and incubated at 37° C for two hours to achieve complete digestion. Following RNAse T2 treatment, the SL pulldowns were each washed five times with 500 µl washing buffer and suspended in 100 µl decapping buffer. The RNA from the U4 and the two SL pulldowns was melted off the bead-bound oligonucleotide at 70° C for 5 min and separated immediately from the beads were immediately removed. 1 µl of each sample was then imaged on the Agilent Bioanalyzer 2100 using the small RNA kit to verify product sizes. 2.5 µl of DCP2 (Enzymax, Lexington, KY) was added to each

SL pulldown and decapping was performed at 37° C for 30 min. The 22- base presumed spliced leader from the large RNA fraction was removed from the excised 5′ cap by addition of 100 µl of decapping buffer containing streptavidin coated beads bound to the spliced leader complementary oligo and annealing of the presumed spliced leader to the oligo at 45° C for 5 min. The beads were removed, suspended in 100 µl of decapping buffer, and the presumed spliced leader was melted off the oligo at 70° C for 5 min and the beads removed. The SL pulldown from the small RNA fraction was transferred to a 3000 NMWL regenerated cellulose Amicon spin filter (Millipore, Billerica, MA). The removed 5′ cap was enriched by increasing the sample volume with 500 µl decapping buffer and passing the sample through the filter at 10,000 x g twice. The resultant samples from the U4 pulldown and the isolated caps and decapped substrates from the two SL pulldowns were lyophilized and prepared for compositional analysis.

**Compositional analysis of purified RNAs**

Each purified RNA sample was aliquoted to a final concentration of 0.73 ng/µl, 7.3 ng in 10 µl of RNAse free water and 10 pg/µl, 100 pg in 10 µl, of isotopically labelled [$^{13}$C][$^{15}$N]-guanosine as internal standard. Each individual sample underwent two-step enzymatic hydrolysis followed by ultra-high performance liquid chromatography (UHPLC) tandem mass spectrometry (MS/MS) method as previously described (ref).[1] The first part of the digestion involves an endonucleolytic cleavage to yield 5′-phosphate with nuclease P1. One unit of nuclease P1 from *Penicillium citrinum* (Sigma-Aldrich) was added to each sample and incubated overnight at 37 ºC. The second step of the hydrolysis was performed by the addition of a unit of bacterial alkaline phosphatase from *E. coli* (Sigma-Aldrich) at 37 ºC for 2 h. This enzyme specifically cleaves the 5′-phosphate from the nucleoside resulting in individual nucleosides and inorganic phosphate. The nucleoside products were lyophilized and reconstituted in 40 µl of RNAse free water (18.0 MΩcm$^{-1}$) containing 0.01 % formic acid prior to UHPLC-MS/MS analysis.

Nucleoside mixture products from hydrolysis were subject to chromatographic separation on a Waters Acquity I-Class UPLC™ (Waters, USA) equipped with a binary pump and auto-sampler maintained at 4 ºC. A Waters Acquity UPLC™ HSS T3 guard column (2.1 x 5 mm 1.8 µm) followed a HSS T3 column (2.1 x 50 mm 1.7 µm). Column temperature was set at 25 ºC. The mobile phases included RNAse-free water (18.0 MΩcm$^{-1}$) containing 0.01% formic acid pH 3.5 (Buffer A) and 50 % acetonitrile in aqueous 0.01 % formic acid (Buffer B). Flow rate was set up at 0.2 ml/min and a gradient applied as described previously.[1]

Tandem MS analysis of nucleosides provides a second dimension analysis whereby the induction of collision energy as the protonated molecular ion [MH$^+$] passes through the collision cell will produce a specific secondary ion or product ion [BH$_2^+$]. Generally, the protonated nucleoside, molecular ion [MH$^+$], is fragmented at the glycosidic bond providing the protonated nucleobase, product ion [BH$_2^+$] and the neutral sugar residue. Tandem MS analysis was performed on a Waters XEVO TQ-S™ (Waters, USA) triple quadrupole mass spectrometer equipped with an electrospray ionization (ESI) source maintained at 150 ºC and the capillary voltage was set at 1 kV. Nitrogen was used as the nebulizer gas which was maintained at 7 bars of pressure, flow rate of 500 l/h and a temperature of 500 ºC. UPLC-MS/MS

analysis was determined in ESI positive-ion and multiple reaction monitoring (MRM) mode with retention times and corresponding molecular and product ion pairs $[MH^+]/[BH_2^+]$ as input parameters.

Quantitation of the 35 nucleosides, 4 majors and 31 modifications, was performed using standard curves with concentration ranging between 0.05 and 100 pg/ul of pure nucleosides and isotope-labelled internal standard $[^{13}C][^{15}N]$-guanosine. Concentrations of nucleosides were calculated following Beer-Lambert law where the extinction coefficient was itself calculated at the absorbance maximum nearest 260 nm rather than as the standard 260 nm as most of the nucleosides have peak maxima that differs from that at 260nm.

To determine the presence of RNA modifications from RNA extracts and to quantify them, we performed UHPLC-MS/MS measurements. A negative control sample was included in the data set to account for background signal from possible modifications part of enzyme's background that could interfere with nucleosides of interest. The negative control contained the enzymes, internal standard and reagents used during the enzymatic digestion. After digestion each sample was lyophilized and reconstituted in RNAse free water (18.0 $M\Omega cm^{-1}$) containing 0.01 % formic acid to a final concentration of 180 pg/µl of RNA and 1 pg/µl if internal standard. A blank sample containing water in 0.01 % in formic acid solution was analyzed between each sample to avoid cross-contamination. Each sample type included 3 biological replicates and 3 technical replicates to account for instrument variability. From the 31 RNA modifications included in the UHPLC-MS/MS method, 10 were detected above their limit of detection, including mostly methylations and pseudouridine. Those modifications below the limit of detection were excluded from the results. Since it was not possible to take total RNA through the streptavidin based enrichment process to use as a negative control, modified RNAs detected in the sample that contained the isolated $m^7G$ cap were used to normalize other samples. Since these other residues are theoretically degraded RNA moieties carried through the enrichment process, they were our best estimate for the levels of modifications in the total RNA pool used, mainly ribosomal and mRNAs.

**Identification of *Amphidinium carterae* cap structure and RNA modifying genes**

The capping on an mRNA or theoretically the spliced leader RNA occurs by three enzymes: an RNA triphosphatase, a guanyl transferase, and a methyl transferase that are found separate or as various fusions when looking at metazoans, fungi, or viruses (Shuman, 2002; Wang, Deng, Ho, & Shuman, 1997). BLAST searches were performed as an early screen to identify contigs in the *A. carterae* transcriptome with the necessary domains to perform the capping reactions. RNA 5'-triphosphatases were common, possibly performing roles in poly-phosphate storage, and further screening was performed to identify open reading frames with the Cet1 domain that had high identity to the *Plasmodium falciparum* 5'-RNA triphosphatase (Gong, Smith, & Shuman, 2006). In addition to methyl transferases involved in capping, additional 2'-Oo-methyl transferases were examined for possible roles in spliced leader modifications similar to those found in trypanosomes (Mittra et al., 2008). Candidate contigs were then used as queries to retrieve similar sequences from other transcriptomes, and a phylogenetic tree was created using these and apicomplexan sequences to verify the reconstruction of the organismal phylogeny.

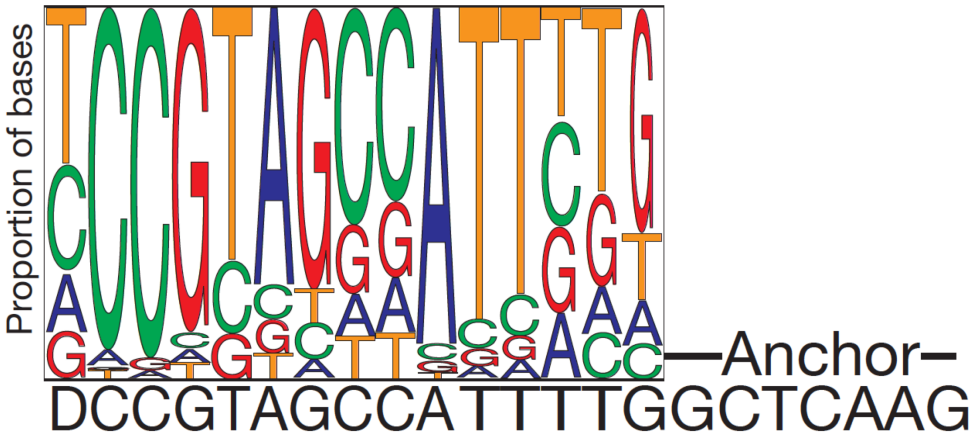



**Figure 1. LOGO diagram of base proportions in the observed spliced leader sequences.**
The relative proportion of nucleotides retrieved within 50 bases of the 5-prime end of dinoflagellate transcripts is shown above the canonical spliced leader sequence from Zhang *et al* 2007. The “anchor” sequence use to retrieve potential spliced leader sequences bioinformatically is shown in the right side consisting of “GCTCAAG”.

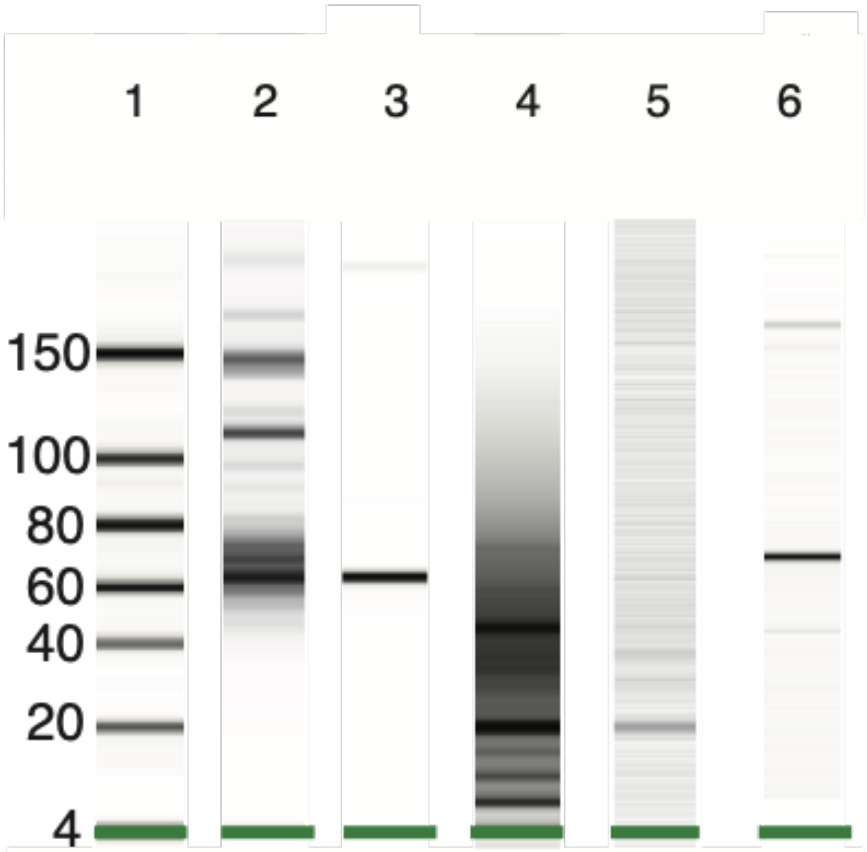



**Figure 2: Products recovered from total small RNA using biotinylated oligonucleotides complementary to the SL and U4 RNAs.**
A virtual gel from an Agilent Bioanalyzer small RNA kit is shown with the size standard in lane **1**. SL (Lane **3**) and U4 (Lane **6**) containing sequences from a small RNA fraction (<200 bases, Lane **2**) were retrieved using biotinylated oligonucleotides complementary to either SL or U4 RNA giving a major band at 64 and 69 bases, respectively. Small RNAs with sequence complementary to the SL were subsequently treated with T2 RNAse (Lane **4**) yielding a 22 base putative SL sequence and subsequent degradation products plus some larger fragments. RNAs from the >200 base fraction were also pulled down with the SL oligonucleotide and treated with RNAse A (Lane **5**) yielding a 22 base putative SL and some minor products at larger sizes.

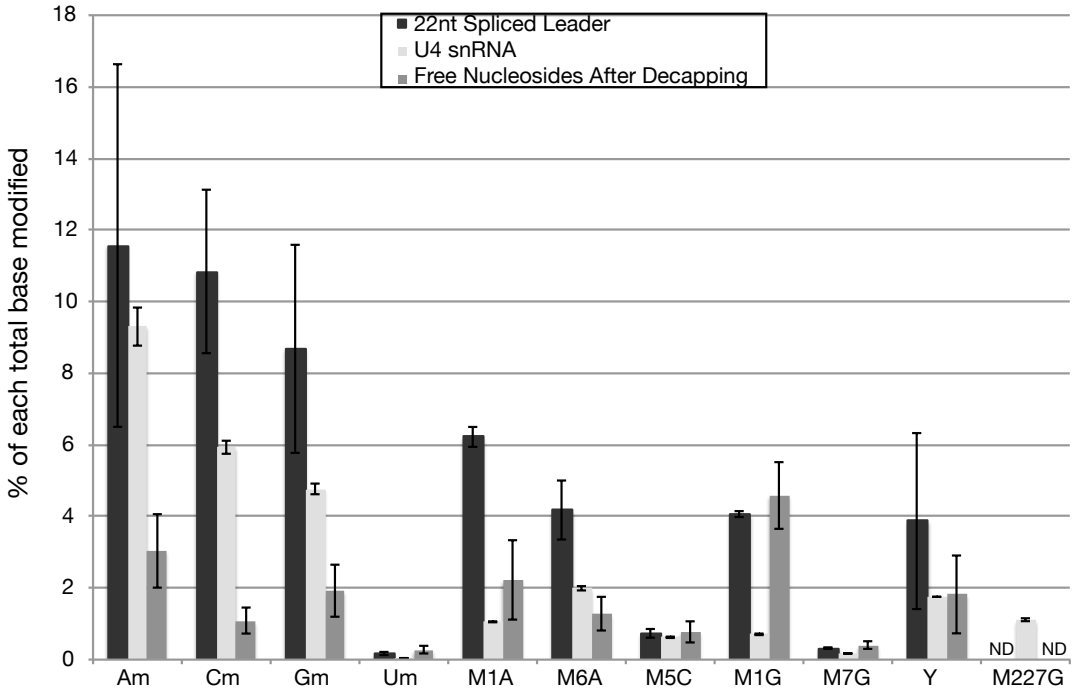



**Figure 3. Modified bases in the 22-nucleotide SL, U4 snRNA, and decapping products.** A graph of the percent of each non-standard ribonucleoside from the total detectable bases is shown with the specific moiety on the X axis and percent on the Y axis or "ND" when not detectable. The first four classes are totals for each non-standard nucleoside followed by specific moieties for which there were standards: 1-methyl adenosine (M1A), 6-methyl adenosine (M6A), 5-methyl cytosine (M5C), 1-methyl guanosine (M1G), 7-methyl guanosine (M7G), pseudouridine (Y), and 2,2,7-trimethyl guanosine (M227G). The RNAse A degradation of spliced leader isolates is shown in black, the U4 snRNA isolate is shown in light grey, and the free nucleosides following decapping of the 22nt spliced leader are shown in dark grey. Error bars represent triplicate compositional analyses from a single sample. Moieties other than 7-methyl guanosine following decapping are likely contaminants from the total RNA pool bound to the Sepharose beads used in sequence enrichment.

# Bibliography


Adolf, J. E., Krupatkina, D., Bachvaroff, T., & Place, A. R. (2007). Karlotoxin mediates grazing by Oxyrrhis marina on strains of Karlodinium veneficum. *Harmful Algae*, *6*(3), 400-412. doi:10.1016/j.hal.2006.12.003

Aitken, C. E., & Lorsch, J. R. (2012). A mechanistic overview of translation initiation in eukaryotes. *Nat Struct Mol Biol*, *19*(6), 568-576. doi:10.1038/nsmb.2303

Akimoto, H., Wu, C., Kinumi, T., & Ohmiya, Y. (2004). Biological rhythmicity in expressed proteins of the marine dinoflagellate Lingulodinium polyedrum demonstrated by chronological proteomics. *Biochem Biophys Res Commun*, *315*(2), 306-312. doi:10.1016/j.bbrc.2004.01.054

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). *Molecular Biology of the Cell*. Garland Science. Retrieved from http://books.google.com/books?id=vBgjzQEACAAJ&hl=&source=gbs_api

Allen, J. R., Roberts, M., Loeblich, A. R., & Klotz, L. C. (1975). Characterization of the DNA from the dinoflagellate Crypthecodinium cohnii and implications for nuclear organization. *Cell*, *6*(2), 161-169. doi:10.1016/0092-8674(75)90006-9

Allender, C. J., LeCleir, G. R., Rinta-Kanto, J. M., Small, R. L., Satchwell, M. F., Boyer, G. L., & Wilhelm, S. W. (2009). Identifying the source of unknown microcystin genes and predicting microcystin variants by comparing genes within uncultured cyanobacterial cells. *Appl Environ Microbiol*, *75*(11), 3598-3604. doi:10.1128/AEM.02448-08

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, *215*(3), 403-410. Retrieved from Google Scholar

Anderson, D. M. (1994). Red tides. *Scientific American*, *271*(2), 62-68. Retrieved from Google Scholar

Angov, E., Hillier, C. J., Kincaid, R. L., & Lyon, J. A. (2008). Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PLoS One*, *3*(5), e2189. doi:10.1371/journal.pone.0002189

Aranda, M., Li, Y., Liew, Y. J., Baumgarten, S., Simakov, O., Wilson, M. C., . . . Voolstra, C. R. (2016). Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. *Sci Rep*, *6*, 39734. doi:10.1038/srep39734

Bachvaroff, T. R. (2019). A precedented nuclear genetic code with all three termination codons reassigned as sense codons in the syndinean Amoebophrya sp. ex Karlodinium veneficum. *PLoS One*, *14*(2), e0212912. doi:10.1371/journal.pone.0212912

Bachvaroff, T. R., Handy, S. M., Place, A. R., & Delwiche, C. F. (2011). Alveolate phylogeny inferred using concatenated ribosomal proteins. *J Eukaryot Microbiol*, *58*(3), 223-233. doi:10.1111/j.1550-7408.2011.00555.x

Bachvaroff, T. R., Kim, S., Guillou, L., Delwiche, C. F., & Coats, D. W. (2012). Molecular diversity of the syndinean genus Euduboscquella based on single-cell PCR analysis. *Appl Environ Microbiol*, *78*(2), 334-345. doi:10.1128/AEM.06678-11

Bachvaroff, T. R., Sanchez Puerta, M. V., & Delwiche, C. F. (2005). Chlorophyll c-containing plastid relationships based on analyses of a multigene data set with all four chromalveolate lineages. *Mol Biol Evol*, *22*(9), 1772-1782. doi:10.1093/molbev/msi172

Bachvaroff, T. R., Gornik, S. G., Concepcion, G. T., Waller, R. F., Mendez, G. S., Lippmeier, J. C., & Delwiche, C. F. (2014). Dinoflagellate phylogeny revisited: using ribosomal proteins to resolve deep branching dinoflagellate clades. *Mol Phylogenet Evol*, *70*, 314-322. doi:10.1016/j.ympev.2013.10.007

Bachvaroff, T. R., & Place, A. R. (2008). From stop to start: tandem gene arrangement, copy number and trans-splicing sites in the dinoflagellate Amphidinium carterae. *PLoS One*, *3*(8), e2929. doi:10.1371/journal.pone.0002929

Bachvaroff, T. R., Place, A. R., & Coats, D. W. (2009). Expressed sequence tags from Amoebophrya sp. infecting Karlodinium veneficum: comparing host and parasite sequences. *J Eukaryot Microbiol*, *56*(6), 531-541. doi:10.1111/j.1550-7408.2009.00433.x

Bachvaroff, T. R., Williams, E. P., Jagus, R., & Place, A. R. (2015). *A cryptic noncanonical multi-module PKS/NRPS found in dinoflagellates*. Proceedings from The 16 International Conference on Harmful Algae, Wellington, New Xealand.

Baden, D. G. (1989). Brevetoxins: unique polyether dinoflagellate toxins. *FASEB J*, *3*(7), 1807-1817.

Baumgarten, S., Cziesielski, M. J., Thomas, L., Michell, C. T., Esherick, L. Y., Pringle, J. R., . . . Voolstra, C. R. (2018). Evidence for miRNA-mediated modulation of the host transcriptome in cnidarian-dinoflagellate symbiosis. *Mol Ecol*, *27*(2), 403-418. doi:10.1111/mec.14452

Beauchemin, M., Roy, S., Daoust, P., Dagenais-Bellefeuille, S., Bertomeu, T., Letourneau, L., . . . Morse, D. (2012). Dinoflagellate tandem array gene transcripts are highly conserved and not polycistronic. *Proc Natl Acad Sci U S A*, *109*(39), 15793-15798. doi:10.1073/pnas.1206683109

Beedessee, G., Hisata, K., Roy, M. C., Van Dolah, F. M., Satoh, N., & Shoguchi, E. (2019). Diversified secondary metabolite biosynthesis gene repertoire revealed in symbiotic dinoflagellates. *Sci Rep*, *9*(1), 1204. doi:10.1038/s41598-018-37792-0

Beedessee, G., Kubota, T., Arimoto, A., Nishitsuji, K., Waller, R. F., Hisata, K., . . . Shoguchi, E. (2020). Integrated omics unveil the secondary metabolic landscape of a basal dinoflagellate. *BMC Biol*, *18*(1), 139. doi:10.1186/s12915-020-00873-6

Beld, J., Sonnenschein, E. C., Vickery, C. R., Noel, J. P., & Burkart, M. D. (2014). The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life. *Nat Prod Rep*, *31*(1), 61-108. doi:10.1039/c3np70054b

Bentley, R., & Bennett, J. W. (1999). Constructing polyketides: from collie to combinatorial biosynthesis. *Annual Reviews in Microbiology*, *53*(1), 411-446. Retrieved from https://www.annualreviews.org/doi/full/10.1146/annurev.micro.53.1.411

Berges, J. A., Franklin, D. J., & Harrison, P. J. (2004). Evolution of an artificial seawater medium: Improvements in enriched seawater, artificial water over the last two decades (Vol. 37: 1138--1145). *J Phycol*, *40*, 619. Retrieved from Google Scholar

Bhaud, Y., Guillebault, D., Lennon, J., Defacque, H., Soyer-Gobillard, M. O., & Moreau, H. (2000). Morphology and behaviour of dinoflagellate chromosomes during the cell cycle and mitosis. *J Cell Sci*, *113*(Pt 7), 1231-1239. doi:10.1242/jcs.113.7.1231

Blanco, A. V., & Chapman, G. B. (1987). Ultrastructural Features of the Marine Dinoflagellate Amphidinium klebsii (Dinophyceae). *Transactions of the American Microscopical Society*, *106*(3), 201. doi:10.2307/3226250

Boere, A. C., Abbas, B., Rijpstra, W. I. C., Versteegh, G. J. M., Volkman, J. K., Sinninghe Damsté, J. S., & Coolen, M. J. L. (2009). Late-Holocene succession of dinoflagellates in an Antarctic fjord using a multi-proxy approach: paleoenvironmental genomics, lipid biomarkers and palynomorphs. *Geobiology*, *7*(3), 265-281. doi:10.1111/j.1472-4669.2009.00202.x

Buhman, K. K., Chen, H. C., & Farese, R. V. (2001). The enzymes of neutral lipid synthesis. *J Biol Chem*, *276*(44), 40369-40372. doi:10.1074/jbc.R100050200

Bunkoczi, G., Pasta, S., Joshi, A., Wu, X., Kavanagh, K. L., Smith, S., & Oppermann, U. (2007). Mechanism and substrate recognition of human holo ACP synthase. *Chem Biol*, *14*(11), 1243-1253. doi:10.1016/j.chembiol.2007.10.013

Cai, X., Herschap, D., & Zhu, G. (2005). Functional characterization of an evolutionarily distinct phosphopantetheinyl transferase in the apicomplexan Cryptosporidium parvum. *Eukaryot Cell*, *4*(7), 1211-1220. doi:10.1128/EC.4.7.1211-1220.2005

Carbonera, D., Di Valentin, M., Spezia, R., & Mezzetti, A. (2014). The unique photophysical properties of the Peridinin-Chlorophyll-α-Protein. *Curr Protein Pept Sci*, *15*(4), 332-350. doi:10.2174/1389203715666140327111139

Cavalier-Smith, T. (1998). A revised six-kingdom system of life. *Biol Rev Camb Philos Soc*, *73*(3), 203-266. doi:10.1017/s0006323198005167

Cavalier-Smith, T. (2002). Chloroplast evolution: secondary symbiogenesis and multiple losses. *Curr Biol*, *12*(2), R62-4. doi:10.1016/s0960-9822(01)00675-3

Cecchin, M., Paloschi, M., Busnardo, G., Cazzaniga, S., Cuine, S., Li-Beisson, Y., . . . Ballottari, M. (2021). $CO_2$ supply modulates lipid remodelling, photosynthetic and respiratory activities in Chlorella species. *Plant Cell Environ*, *44*(9), 2987-3001. doi:10.1111/pce.14074

Chan, C. X., Baglivi, F. L., Jenkins, C. E., & Bhattacharya, D. (2013). Foreign gene recruitment to the fatty acid biosynthesis pathway in diatoms. *Mob Genet Elements*, *3*(5), e27313. doi:10.4161/mge.27313

Chen, W., Colon, R., Louda, J. W., Del Rey, F. R., Durham, M., & Rein, K. S. (2018). Brevetoxin (PbTx-2) influences the redox status and NPQ of Karenia

brevis by way of thioredoxin reductase. *Harmful Algae*, *71*, 29-39. doi:10.1016/j.hal.2017.11.004
Chow, M. H., Yan, K. T., Bennett, M. J., & Wong, J. T. (2010). Birefringence and DNA condensation of liquid crystalline chromosomes. *Eukaryot Cell*, *9*(10), 1577-1587. doi:10.1128/EC.00026-10
Clairfeuille, T., Norwood, S. J., Qi, X., Teasdale, R. D., & Collins, B. M. (2015). Structure and Membrane Binding Properties of the Endosomal Tetratricopeptide Repeat (TPR) Domain-containing Sorting Nexins SNX20 and SNX21. *J Biol Chem*, *290*(23), 14504-14517. doi:10.1074/jbc.M115.650598
Clayton, C. (2019). Regulation of gene expression in trypanosomatids: living with polycistronic transcription. *Open Biol*, *9*(6), 190072. doi:10.1098/rsob.190072
Cloern, J. E., Foster, S. Q., & Kleckner, A. E. (2014). Phytoplankton primary production in the world's estuarine-coastal ecosystems. *Biogeosciences*, *11*(9), 2477-2501. doi:10.5194/bg-11-2477-2014
Coats, D. W., & Park, M. G. (2002). PARASITISM OF PHOTOSYNTHETIC DINOFLAGELLATES BY THREE STRAINS OF AMOEBOPHRYA (DINOPHYTA): PARASITE SURVIVAL, INFECTIVITY, GENERATION TIME, AND HOST SPECIFICITY 1. *Journal of Phycology*, *38*(3), 520-528. doi:10.1046/j.1529-8817.2002.01200.x
Coats, D. W., Tyler, M. A., & Anderson, D. M. (1984). SEXUAL PROCESSES IN THE LIFE CYCLE OFGYRODINIUM UNCATENUM(DINOPHYCEAE): A MORPHOGENETIC OVERVIEW1. *Journal of Phycology*, *20*(3), 351-361. doi:10.1111/j.0022-3646.1984.00351.x
Coats, D. W., Bachvaroff, T. R., & Delwiche, C. F. (2012). Revision of the family Duboscquellidae with description of Euduboscquella crenulata n. gen., n. sp. (Dinoflagellata, Syndinea), an intracellular parasite of the ciliate Favella panamensis Kofoid & Campbell. *J Eukaryot Microbiol*, *59*(1), 1-11. doi:10.1111/j.1550-7408.2011.00588.x
Cointet, E., Wielgosz-Collin, G., Bougaran, G., Rabesaotra, V., Gonçalves, O., & Méléder, V. (2019). Effects of light and nitrogen availability on photosynthetic efficiency and fatty acid content of three original benthic diatom strains. *PLoS One*, *14*(11), e0224701. doi:10.1371/journal.pone.0224701
Colon, R., Wheater, M., Joyce, E. J., Ste Marie, E. J., Hondal, R. J., & Rein, K. S. (2021). The Marine Neurotoxin Brevetoxin (PbTx-2) Inhibits Karenia brevis and Mammalian Thioredoxin Reductases by Targeting Different Residues. *J Nat Prod*, *84*(11), 2961-2970. doi:10.1021/acs.jnatprod.1c00795
Costas, E., & Goyanes, V. (2005). Architecture and evolution of dinoflagellate chromosomes: an enigmatic origin. *Cytogenet Genome Res*, *109*(1-3), 268-275. doi:10.1159/000082409
Dagenais-Bellefeuille, S., Beauchemin, M., & Morse, D. (2017). miRNAs Do Not Regulate Circadian Protein Synthesis in the Dinoflagellate Lingulodinium polyedrum. *PLoS One*, *12*(1), e0168817. doi:10.1371/journal.pone.0168817
Dagenais-Bellefeuille, S., & Morse, D. (2013). Putting the N in dinoflagellates. *Front Microbiol*, *4*, 369. doi:10.3389/fmicb.2013.00369

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., . . . Karsenti, E. (2015). Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean. *Science*, *348*(6237), 1261605. doi:10.1126/science.1261605

Decroly, E., Ferron, F., Lescar, J., & Canard, B. (2011). Conventional and unconventional mechanisms for capping viral mRNA. *Nat Rev Microbiol*, *10*(1), 51-65. doi:10.1038/nrmicro2675

Deeds, J. R., Terlizzi, D. E., Adolf, J. E., Stoecker, D. K., & Place, A. R. (2002). Toxic activity from cultures of Karlodinium micrum (=Gyrodinium galatheanum) (Dinophyceae)—a dinoflagellate associated with fish mortalities in an estuarine aquaculture facility. *Harmful Algae*, *1*(2), 169-189. doi:10.1016/s1568-9883(02)00027-6

Delwiche. (1999). Tracing the Thread of Plastid Diversity through the Tapestry of Life. *Am Nat*, *154*(S4), S164-S177. doi:10.1086/303291

Desjardins, M., & Morse, D. (1993). The polypeptide components of scintillons, the bioluminescence organelles of the dinoflagellate Gonyaulax polyedra. *Biochem Cell Biol*, *71*(3-4), 176-182. doi:10.1139/o93-028

Diao, J., Song, X., Zhang, X., Chen, L., & Zhang, W. (2018). Genetic Engineering of Crypthecodinium cohnii to Increase Growth and Lipid Accumulation. *Front Microbiol*, *9*, 492. doi:10.3389/fmicb.2018.00492

DODGE, J. D., & CRAWFORD, R. M. (1970). A survey of thecal fine structure in the Dinophyceae. *Botanical Journal of the Linnean Society*, *63*(1), 53-67. doi:10.1111/j.1095-8339.1970.tb02302.x

Dodge, J. D. (1995). Thecal structure, taxonomy, and distribution of the planktonic dinoflagellate Micracanthodinium setiferum (Gonyaulacales, Dinophyceae). *Phycologia*, *34*(4), 307-312. doi:10.2216/i0031-8884-34-4-307.1

Dong, H. P., Williams, E., Wang, D. Z., Xie, Z. X., Hsia, R. C., Jenck, A., . . . Place, A. R. (2013). Responses of Nannochloropsis oceanica IMET1 to Long-Term Nitrogen Starvation and Recovery. *Plant Physiol*, *162*(2), 1110-1126. doi:10.1104/pp.113.214320

Dorrell, R. G., & Howe, C. J. (2015). Integration of plastids with their hosts: Lessons learned from dinoflagellates. *Proc Natl Acad Sci U S A*, *112*(33), 10247-10254. doi:10.1073/pnas.1421380112

Du, L., Sánchez, C., & Shen, B. (2001). Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules. *Metab Eng*, *3*(1), 78-95. doi:10.1006/mben.2000.0171

Du, Q., Sui, Z., Chang, L., Wei, H., Liu, Y., Mi, P., . . . Que, Z. (2016). Genome size of Alexandrium catenella and Gracilariopsis lemaneiformis estimated by flow cytometry. *Journal of Ocean University of China*, *15*(4), 704-710. doi:10.1007/s11802-016-2988-7

Dunstan, G. A., Volkman, J. K., Barrett, S. M., Leroi, J.-M., & Jeffrey, S. W. (1993). Essential polyunsaturated fatty acids from 14 species of diatom (Bacillariophyceae). *Phytochemistry*, *35*(1), 155-161. doi:10.1016/s0031-9422(00)90525-9

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, *32*(5), 1792-1797. doi:10.1093/nar/gkh340

Evitt, W. R. (1961). Observations on the Morphology of Fossil Dinoflagellates. *Micropaleontology*, *7*(4), 385. doi:10.2307/1484378

Fagan, T., Morse, D., & Hastings, J. W. (1999). Circadian synthesis of a nuclear-encoded chloroplast glyceraldehyde-3-phosphate dehydrogenase in the dinoflagellate Gonyaulax polyedra is translationally controlled. *Biochemistry*, *38*(24), 7689-7695. doi:10.1021/bi9826005

Favrot, L., Blanchard, J. S., & Vergnolle, O. (2016). Bacterial GCN5-RelatedN-Acetyltransferases: From Resistance to Regulation. *Biochemistry*, *55*(7), 989-1002. doi:10.1021/acs.biochem.5b01269

Felnagle, E. A., Jackson, E. E., Chan, Y. A., Podevels, A. M., Berti, A. D., McMahon, M. D., & Thomas, M. G. (2008). Nonribosomal peptide synthetases involved in the production of medically relevant natural products. *Mol Pharm*, *5*(2), 191-211. doi:10.1021/mp700137g

Fewer, D. P., Rouhiainen, L., Jokela, J., Wahlsten, M., Laakso, K., Wang, H., & Sivonen, K. (2007). Recurrent adenylation domain replacement in the microcystin synthetase gene cluster. *BMC Evol Biol*, *7*, 183. doi:10.1186/1471-2148-7-183

Fleming, A. (2001). On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. 1929. *Bull World Health Organ*, *79*(8), 780-790. Retrieved from https://pubmed.ncbi.nlm.nih.gov/11545337

Flugel, R. S., Hwangbo, Y., Lambalot, R. H., Cronan, J. E., & Walsh, C. T. (2000). Holo-(acyl carrier protein) synthase and phosphopantetheinyl transfer in Escherichia coli. *J Biol Chem*, *275*(2), 959-968. doi:10.1074/jbc.275.2.959

Franke, J., Ishida, K., & Hertweck, C. (2012). Genomics-driven discovery of burkholderic acid, a noncanonical, cryptic polyketide from human pathogenic Burkholderia species. *Angew Chem Int Ed Engl*, *51*(46), 11611-11615. doi:10.1002/anie.201205566

Frickey, T., & Lupas, A. (2004). CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics*, *20*(18), 3702-3704. doi:10.1093/bioinformatics/bth444

Fukatsu, T., Onodera, K., Ohta, Y., Oba, Y., Nakamura, H., Shintani, T., . . . Ojika, M. (2007). Zooxanthellamide D, a polyhydroxy polyene amide from a marine dinoflagellate, and chemotaxonomic perspective of the Symbiodinium polyols. *J Nat Prod*, *70*(3), 407-411. doi:10.1021/np060596p

Fukuda, Y., & Suzaki, T. (2015). Unusual Features of Dinokaryon, the Enigmatic Nucleus of Dinoflagellates. In *Marine Protists* (pp. 23-45). Tokyo: Springer Japan. doi:10.1007/978-4-431-55130-0_2

Gajadhar, A. A., Marquardt, W. C., Hall, R., Gunderson, J., Ariztia-Carmona, E. V., & Sogin, M. L. (1991). Ribosomal RNA sequences of Sarcocystis muris, Theileria annulata and Crypthecodinium cohnii reveal evolutionary relationships among apicomplexans, dinoflagellates, and ciliates. *Mol Biochem Parasitol*, *45*(1), 147-154. doi:10.1016/0166-6851(91)90036-6

Galleron, C. (1976). SYNCHRONIZATION OF THE MARINE DINOFLAGELLATE AMPHIDINIUM CARTERI IN DENSE

CULTURES<sup>1</sup>. *Journal of Phycology*, *12*(1), 69-73. doi:10.1111/j.1529-8817.1976.tb02828.x
Gao, D., Qiu, L., Hou, Z., Zhang, Q., Wu, J., Gao, Q., & Song, L. (2013). Computational Identification of MicroRNAs from the Expressed Sequence Tags of Toxic Dinoflagellate Alexandrium Tamarense. *Evol Bioinform Online*, *9*, 479-485. doi:10.4137/EBO.S12899
Gao, X. P., & Li, J. Y. (1986). Nuclear division in the marine dinoflagellate Oxyrrhis marina. *J Cell Sci*, *85*, 161-175. doi:10.1242/jcs.85.1.161
Gavelis, G. S., Herranz, M., Wakeman, K. C., Ripken, C., Mitarai, S., Gile, G. H., . . . Leander, B. S. (2019). Dinoflagellate nucleus contains an extensive endomembrane network, the nuclear net. *Sci Rep*, *9*(1), 839. doi:10.1038/s41598-018-37065-w
Geerlof, A., Lewendon, A., & Shaw, W. V. (1999). Purification and Characterization of Phosphopantetheine Adenylyltransferase fromEscherichia coli. *Journal of Biological Chemistry*, *274*(38), 27105-27111. doi:10.1074/jbc.274.38.27105
Gerc, A. J., Stanley-Wall, N. R., & Coulthurst, S. J. (2014). Role of the phosphopantetheinyltransferase enzyme, PswP, in the biosynthesis of antimicrobial secondary metabolites by Serratia marcescens Db10. *Microbiology (Reading)*, *160*(Pt 8), 1609-1617. doi:10.1099/mic.0.078576-0
Gomes, E. S., Schuch, V., & de Macedo Lemos, E. G. (2013). Biotechnology of polyketides: new breath of life for the novel antibiotic genetic pathways discovery through metagenomics. *Braz J Microbiol*, *44*(4), 1007-1034. doi:10.1590/s1517-83822013000400002
Gómez, F., & Artigas, L. F. (2019). Redefinition of the Dinoflagellate GenusAlexandriumBased onCentrodinium: Reinstatement ofGessneriumandProtogonyaulax, andEpisemicolongen. nov. (Gonyaulacales, Dinophyceae). *Journal of Marine Biology*, *2019*, 1-17. doi:10.1155/2019/1284104
Gong, C., Smith, P., & Shuman, S. (2006). Structure-function analysis of Plasmodium RNA triphosphatase and description of a triphosphate tunnel metalloenzyme superfamily that includes Cet1-like RNA triphosphatases and CYTH proteins. *RNA*, *12*(8), 1468-1474. doi:10.1261/rna.119806
Gornik, S. G., Cassin, A. M., MacRae, J. I., Ramaprasad, A., Rchiad, Z., McConville, M. J., . . . Waller, R. F. (2015). Endosymbiosis undone by stepwise elimination of the plastid in a parasitic dinoflagellate. *Proceedings of the National Academy of Sciences*, *112*(18), 5767-5772. Retrieved from https://www.pnas.org/content/112/18/5767.short
Gornik, S. G., Ford, K. L., Mulhern, T. D., Bacic, A., McFadden, G. I., & Waller, R. F. (2012). Loss of nucleosomal DNA condensation coincides with appearance of a novel nuclear protein in dinoflagellates. *Curr Biol*, *22*(24), 2303-2312. doi:10.1016/j.cub.2012.10.036
Gornik, S. G., Hu, I., Lassadi, I., & Waller, R. F. (2019). The Biochemistry and Evolution of the Dinoflagellate Nucleus. *Microorganisms*, *7*(8). doi:10.3390/microorganisms7080245
Guihéneuf, F., & Stengel, D. B. (2013). LC-PUFA-enriched oil production by microalgae: accumulation of lipid and triacylglycerols containing n-3 LC-PUFA

is triggered by nitrogen limitation and inorganic carbon availability in the marine haptophyte Pavlova lutheri. *Mar Drugs*, *11*, 4246-4266. doi:10.3390/md11114246

Guo, X., Wang, Z., Liu, L., & Li, Y. (2021). Transcriptome and metabolome analyses of cold and darkness-induced pellicle cysts of Scrippsiella trochoidea. *BMC Genomics*, *22*(1), 526. doi:10.1186/s12864-021-07840-7

Gurney, R., & Thomas, C. M. (2011). Mupirocin: biosynthesis, special features and applications of an antibiotic from a gram-negative bacterium. *Appl Microbiol Biotechnol*, *90*(1), 11-21. doi:10.1007/s00253-011-3128-3

Hansen, P. J., Skovgaard, A., & Stoecker, R. N. G. A. D. K. (2000).Physiology of the mixotrophic dinoflagellate Fragilidium subglobosum. *Marine Ecology Progress Series*, *201*(August 9 2000), 137-146. doi:10.2307/24863575

Haq, S., Bachvaroff, T. R., & Place, A. R. (2017). Characterization of Acetyl-CoA Carboxylases in the Basal Dinoflagellate Amphidinium carterae. *Mar Drugs*, *15*(6). doi:10.3390/md15060149

Harada, A., Ohtsuka, S., & Horiguchi, T. (2007). Species of the parasitic genus Duboscquella are members of the enigmatic Marine Alveolate Group I. *Protist*, *158*(3), 337-347. doi:10.1016/j.protis.2007.03.005

Hardeland, R., & Nord, P. (1984). Visualization of free-running circadian rhythms in the dinoflagellatePyrocystis noctiluca. *Marine Behaviour and Physiology*, *11*(3), 199-207. doi:10.1080/10236248409387045

Hastings, J. W. (2013). Circadian Rhythms in Dinoflagellates: What Is the Purpose of Synthesis and Destruction of Proteins. *Microorganisms*, *1*, 26-32. doi:10.3390/microorganisms1010026

Hehenberger, E., Gast, R. J., & Keeling, P. J. (2019). A kleptoplastidic dinoflagellate and the tipping point between transient and fully integrated plastid endosymbiosis. *Proc Natl Acad Sci U S A*, *116*(36), 17934-17942. doi:10.1073/pnas.1910121116

Hershberg, R., & Petrov, D. A. (2008). Selection on codon bias. *Annu Rev Genet*, *42*, 287-299. doi:10.1146/annurev.genet.42.110807.091442

Hertweck, C., Luzhetskyy, A., Rebets, Y., & Bechthold, A. (2007). Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat Prod Rep*, *24*(1), 162-190. doi:10.1039/b507395m

Holm-Hansen, O. (1969). Algae: amounts of DNA and organic carbon in single cells. *Science*, *163*(3862), 87-88. doi:10.1126/science.163.3862.87

Hölzl, G., & Dörmann, P. (2019). Chloroplast Lipids and Their Biosynthesis. *Annu Rev Plant Biol*, *70*, 51-81. doi:10.1146/annurev-arplant-050718-100202

Hong, H.-H., Lee, H.-G., Jo, J., Kim, H. M., Kim, S.-M., Park, J. Y., . . . Kim, K. Y. (2016). The exceptionally large genome of the harmful red tide dinoflagellate Cochlodinium polykrikoides Margalef (Dinophyceae): determination by flow cytometry. *ALGAE*, *31*(4), 373-378. doi:10.4490/algae.2016.31.12.6

Hoppenrath, M., & Leander, B. S. (2010). Dinoflagellate phylogeny as inferred from heat shock protein 90 and ribosomal gene sequences. *PLoS One*, *5*(10), e13220. doi:10.1371/journal.pone.0013220

Hou, Y., & Lin, S. (2009). Distinct gene number-genome size relationships for eukaryotes and non-eukaryotes: gene content estimation for dinoflagellate genomes. *PLoS One*, *4*(9), e6978. doi:10.1371/journal.pone.0006978

Houdai, T., Matsuoka, S., Murata, M., Satake, M., Ota, S., Oshima, Y., & Rhodes, L. L. (2001). Acetate labeling patterns of dinoflagellate polyketides, amphidinols 2, 3 and 4. *Tetrahedron*, *57*(26), 5551-5555. Retrieved from http://scholar.google.com/scholar?output=instlink&nossl=1&q=info:m48ihm3SBNMJ:scholar.google.com/&hl=en&as_sdt=0,21&as_ylo=2001&as_yhi=2001&scillfp=12010515443712812409&oi=lle

Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., . . . Yeats, C. (2009). InterPro: the integrative protein signature database. *Nucleic Acids Res*, *37*(Database issue), D211-5. doi:10.1093/nar/gkn785

Hutcheon, C., Ditt, R. F., Beilstein, M., Comai, L., Schroeder, J., Goldstein, E., . . . Kiser, J. (2010). Polyploid genome of Camelina sativa revealed by isolation of fatty acid synthesis genes. *BMC Plant Biol*, *10*, 233. doi:10.1186/1471-2229-10-233

Ishida, H., Nozawa, A., Totoribe, K., Muramatsu, N., Nukaya, H., Tsuji, K., . . . Berkett, N. (1995). Brevetoxin B1, a new polyether marine toxin from the New Zealand shellfish, Austrovenus stutchburyi. *Tetrahedron letters*, *36*(5), 725-728. Retrieved from http://gateway.webofknowledge.com/gateway/Gateway.cgi?GWVersion=2&SrcApp=GSSearch&SrcAuth=Scholar&DestApp=WOS_CPL&DestLinkType=CitingArticles&UT=A1995QF02200023&SrcURL=https://scholar.google.com/&SrcDesc=Back+to+Google+Scholar&GSPage=TC

Ishida, K.-i., & Green, B. R. (2002). Second- and third-hand chloroplasts in dinoflagellates: phylogeny of oxygen-evolving enhancer 1 (PsbO) protein reveals replacement of a nuclear-encoded plastid gene by that of a haptophyte tertiary endosymbiont. *Proc Natl Acad Sci U S A*, *99*(14), 9294-9299. doi:10.1073/pnas.142091799

Iwataki, M. (2008). Taxonomy and identification of the armored dinoflagellate genus Heterocapsa (Peridiniales, Dinophyceae). *Plankton and Benthos Research*, *3*, 135-142. doi:10.3800/pbr.3.135

Izoré, T., & Cryle, M. J. (2018). The many faces and important roles of protein-protein interactions during non-ribosomal peptide synthesis. *Nat Prod Rep*, *35*(11), 1120-1139. doi:10.1039/c8np00038g

Jacobson, D. M., & Anderson, D. M. (1996). WIDESPREAD PHAGOCYTOSIS OF CILIATES AND OTHER PROTISTS BY MARINE MIXOTROPHIC AND HETEROTROPHIC THECATE DINOFLAGELLATES1. *Journal of Phycology*, *32*(2), 279-285. doi:10.1111/j.0022-3646.1996.00279.x

Janouškovec, J., Gavelis, G. S., Burki, F., Dinh, D., Bachvaroff, T. R., Gornik, S. G., . . . Saldarriaga, J. F. (2017). Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics. *Proc Natl Acad Sci U S A*, *114*(2), E171-E180. doi:10.1073/pnas.1614842114

Janouskovec, J., Horák, A., Oborník, M., Lukes, J., & Keeling, P. J. (2010). A common red algal origin of the apicomplexan, dinoflagellate, and heterokont

plastids. *Proc Natl Acad Sci U S A*, *107*(24), 10949-10954. doi:10.1073/pnas.1003335107

Javed, F., Qadir, M. I., Janbaz, K. H., & Ali, M. (2011). Novel drugs from marine microorganisms. *Crit Rev Microbiol*, *37*(3), 245-249. doi:10.3109/1040841X.2011.576234

Jenke-Kodama, H., & Dittmann, E. (2009). Evolution of metabolic diversity: insights from microbial polyketide synthases. *Phytochemistry*, *70*(15-16), 1858-1866. doi:10.1016/j.phytochem.2009.05.021

Jensen, P. R. (2016). Natural Products and the Gene Cluster Revolution. *Trends Microbiol*, *24*(12), 968-977. doi:10.1016/j.tim.2016.07.006

Jeong, H. J., Yoo, Y. D., Kim, J. S., Seong, K. A., Kang, N. S., & Kim, T. H. (2010). Growth, feeding and ecological roles of the mixotrophic and heterotrophic dinoflagellates in marine planktonic food webs. *Ocean Science Journal*, *45*(2), 65-91. doi:10.1007/s12601-010-0007-2

Jia, Y., Gao, H., Tong, M., & Anderson, D. M. (2019). Cell cycle regulation of the mixotrophic dinoflagellate Dinophysis acuminata: Growth, photosynthetic efficiency and toxin production. *Harmful Algae*, *89*, 101672. doi:10.1016/j.hal.2019.101672

John, U., Beszteri, B., Derelle, E., Van de Peer, Y., Read, B., Moreau, H., & Cembella, A. (2008). Novel insights into evolution of protistan polyketide synthases through phylogenomic analysis. *Protist*, *159*(1), 21-30. doi:10.1016/j.protis.2007.08.001

John, U., Lu, Y., Wohlrab, S., Groth, M., Janouškovec, J., Kohli, G. S., . . . Glöckner, G. (2019). An aerobic eukaryotic parasite with functional mitochondria that likely lacks a mitochondrial genome. *Sci Adv*, *5*(4), eaav1110. doi:10.1126/sciadv.aav1110

Jones, G. D., Williams, E. P., Place, A. R., Jagus, R., & Bachvaroff, T. R. (2015). The alveolate translation initiation factor 4E family reveals a custom toolkit for translational control in core dinoflagellates. *BMC Evol Biol*, *15*, 14. doi:10.1186/s12862-015-0301-9

Kato, K. H., Moriyama, A., Huitorel, P., Cosson, J., Cachon, M., & Sato, H. (1997). Isolation of the major basic nuclear protein and its localization on chromosomes of the dinoflagellate, Oxyrrhis marina. *Biol Cell*, *89*(1), 43-52. doi:10.1016/s0248-4900(99)80080-x

Keeling, P. J. (2010). The endosymbiotic origin, diversification and fate of plastids. *Philos Trans R Soc Lond B Biol Sci*, *365*(1541), 729-748. doi:10.1098/rstb.2009.0103

Kevany, B. M., Rasko, D. A., & Thomas, M. G. (2009). Characterization of the complete zwittermicin A biosynthesis gene cluster from Bacillus cereus. *Appl Environ Microbiol*, *75*(4), 1144-1155. doi:10.1128/AEM.02518-08

Khosla, C. (2009). Structures and mechanisms of polyketide synthases. *J Org Chem*, *74*(17), 6416-6420. doi:10.1021/jo9012089

Khosla, C., Kapur, S., & Cane, D. E. (2009). Revisiting the modularity of modular polyketide synthases. *Curr Opin Chem Biol*, *13*(2), 135-143. doi:10.1016/j.cbpa.2008.12.018

Kobayashi, J. (2008). Amphidinolides and its related macrolides from marine dinoflagellates. *J Antibiot (Tokyo)*, *61*(5), 271-284. doi:10.1038/ja.2008.39

Kohli, G. S., Campbell, K., John, U., Smith, K. F., Fraga, S., Rhodes, L. L., & Murray, S. A. (2017). Role of Modular Polyketide Synthases in the Production of Polyether Ladder Compounds in Ciguatoxin-Producing Gambierdiscus polynesiensis and G. excentricus (Dinophyceae). *J Eukaryot Microbiol*, *64*(5), 691-706. doi:10.1111/jeu.12405

Kohli, G. S., John, U., Figueroa, R. I., Rhodes, L. L., Harwood, D. T., Groth, M., . . . Murray, S. A. (2015). Polyketide synthesis genes associated with toxin production in two species of Gambierdiscus (Dinophyceae). *BMC Genomics*, *16*, 410. doi:10.1186/s12864-015-1625-y

Kohli, G. S., John, U., Van Dolah, F. M., & Murray, S. A. (2016). Evolutionary distinctiveness of fatty acid and polyketide synthesis in eukaryotes. *ISME J*, *10*(8), 1877-1890. doi:10.1038/ismej.2015.263

Korman, T. P., Ames, B., & (Sheryl) Tsai, S.-C. (2010). Structural Enzymology of Polyketide Synthase: The Structure–Sequence–Function Correlation. In *Comprehensive Natural Products II* (pp. 305-345). Elsevier. doi:10.1016/b978-008045382-8.00020-4

LaJeunesse, T. C., Lambert, G., Andersen, R. A., Coffroth, M. A., & Galbraith, D. W. (2005a). SYMBIODINIUM (PYRRHOPHYTA) GENOME SIZES (DNA CONTENT) ARE SMALLEST AMONG DINOFLAGELLATES1. *Journal of Phycology*, *41*(4), 880-886. Retrieved from Google Scholar

LaJeunesse, T. C., Lambert, G., Andersen, R. A., Coffroth, M. A., & Galbraith, D. W. (2005b). SYMBIODINIUM (PYRRHOPHYTA) GENOME SIZES (DNA CONTENT) ARE SMALLEST AMONG DINOFLAGELLATES1. *Journal of Phycology*, *41*(4), 880-886. doi:10.1111/j.0022-3646.2005.04231.x

Lambalot, R. H., Gehring, A. M., Flugel, R. S., Zuber, P., LaCelle, M., Marahiel, M. A., . . . Walsh, C. T. (1996). A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem Biol*, *3*(11), 923-936. doi:10.1016/s1074-5521(96)90181-7

Lapointe, M., & Morse, D. (2008). Reassessing the role of a 3'-UTR-binding translational inhibitor in regulation of circadian bioluminescence rhythm in the dinoflagellate Gonyaulax. *Biol Chem*, *389*(1), 13-19. doi:10.1515/BC.2008.003

Lasda, E. L., & Blumenthal, T. (2011). Trans-splicing. *Wiley Interdiscip Rev RNA*, *2*(3), 417-434. doi:10.1002/wrna.71

Lauritano, C., De Luca, D., Ferrarini, A., Avanzato, C., Minio, A., Esposito, F., & Ianora, A. (2017). De novo transcriptome of the cosmopolitan dinoflagellate Amphidinium carterae to identify enzymes with biotechnological potential. *Sci Rep*, *7*(1), 11701. doi:10.1038/s41598-017-12092-1

Leblond, J. D., Sengco, M. R., Sickman, J. O., Dahmen, J. L., & Anderson, D. M. (2006). Sterols of the syndinian dinoflagellate Amoebophrya sp., a parasite of the dinoflagellate Alexandrium tamarense (Dinophyceae). *J Eukaryot Microbiol*, *53*(3), 211-216. doi:10.1111/j.1550-7408.2006.00097.x

Leblond, J. D., & Dahmen, J. L. (2012). Mono- and digalactosyldiacylglycerol composition of dinoflagellates. V. The galactolipid profile ofAlexandrium tamarense(Dinophyceae) during the course of infection by the parasitic

syndinian dinoflagellateAmoebophryasp. *European Journal of Phycology*, *47*(4), 490-497. doi:10.1080/09670262.2012.742140
Leblond, J. D., Evans, T. J., & Chapman, P. J. (2003). The biochemistry of dinoflagellate lipids, with particular reference to the fatty acid and sterol composition of a Karenia brevis bloom. *Phycologia*, *42*(4), 324-331. doi:10.2216/i0031-8884-42-4-324.1
Lee, J. L.-N., Chiang, K.-P., & Tsai, S.-F. (2021). Sexual Reproduction in Dinoflagellates—The Case of Noctiluca scintillans and Its Ecological Implications. *Frontiers in Marine Science*, *8*. doi:10.3389/fmars.2021.704398
Lee, M. S., Qin, G., Nakanishi, K., & Zagorski, M. G. (1989). Biosynthetic studies of brevetoxins, potent neurotoxins produced by the dinoflagellate Gymnodinium breve. *Journal of the American Chemical Society*, *111*(16), 6234-6241. doi:10.1021/ja00198a039
Levi-Setti, R., Gavrilov, K. L., & Rizzo, P. J. (2008). Divalent cation distribution in dinoflagellate chromosomes imaged by high-resolution ion probe mass spectrometry. *Eur J Cell Biol*, *87*(12), 963-976. doi:10.1016/j.ejcb.2008.06.002
Li, L., & Hastings, J. W. (1998). The structure and organization of the luciferase gene in the photosynthetic dinoflagellate Gonyaulax polyedra. *Plant Mol Biol*, *36*(2), 275-284. doi:10.1023/a:1005941421474
Li, Y., Han, D., Sommerfeld, M., & Hu, Q. (2011). Photosynthetic carbon partitioning and lipid production in the oleaginous microalga Pseudochlorococcum sp. (Chlorophyceae) under nitrogen-limited conditions. *Bioresour Technol*, *102*(1), 123-129. doi:10.1016/j.biortech.2010.06.036
Lidie, K. B., Ryan, J. C., Barbier, M., & Van Dolah, F. M. (2005). Gene expression in Florida red tide dinoflagellate Karenia brevis: analysis of an expressed sequence tag library and development of DNA microarray. *Marine Biotechnology*, *7*(5), 481-493. Retrieved from Google Scholar
Lidie, K. B., & van Dolah, F. M. (2007). Spliced leader RNA-mediated trans-splicing in a dinoflagellate, Karenia brevis. *J Eukaryot Microbiol*, *54*(5), 427-435. doi:10.1111/j.1550-7408.2007.00282.x
Lim, S. K., Ju, J., Zazopoulos, E., Jiang, H., Seo, J. W., Chen, Y., . . . Shen, B. (2009). iso-Migrastatin, migrastatin, and dorrigocin production in Streptomyces platensis NRRL 18993 is governed by a single biosynthetic machinery featuring an acyltransferase-less type I polyketide synthase. *J Biol Chem*, *284*(43), 29746-29756. doi:10.1074/jbc.M109.046805
Lin, S., Van Lanen, S. G., & Shen, B. (2009). A free-standing condensation enzyme catalyzing ester bond formation in C-1027 biosynthesis. *Proc Natl Acad Sci U S A*, *106*(11), 4183-4188. doi:10.1073/pnas.0808880106
Lin, S., Cheng, S., Song, B., Zhong, X., Lin, X., Li, W., . . . Morse, D. (2015). The Symbiodinium kawagutii genome illuminates dinoflagellate gene expression and coral symbiosis. *Science*, *350*(6261), 691-694. doi:10.1126/science.aad0408
Litaker, R. W., Vandersea, M. W., Kibler, S. R., Madden, V. J., Noga, E. J., & Tester, P. A. (2002). Life cycle of the heterotrophic dinoflagellate Pfiesteria piscicida (Dinophyceae). *Journal of Phycology*, *38*(3), 442-463. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1529-8817.2002.01242.x

Liu, C. L., Place, A. R., & Jagus, R. (2017). Use of Antibiotics for Maintenance of Axenic Cultures of Amphidinium carterae for the Analysis of Translation. *Mar Drugs*, *15*(8), E242. doi:10.3390/md15080242

Liu, Z., Zang, X., Cao, X., Wang, Z., Liu, C., Sun, D., . . . Pang, C. (2018). Cloning of the pks3 gene of Aurantiochytrium limacinum and functional study of the 3-ketoacyl-ACP reductase and dehydratase enzyme domains. *PLoS One*, *13*(12), e0208853. doi:10.1371/journal.pone.0208853

Livolant, F. (1978). Positive and negative birefringence in chromosomes. *Chromosoma*, *68*(1), 45-58. doi:10.1007/BF00330371

Lupien, L. E., Dunkley, E. M., Maloy, M. J., Lehner, I. B., Foisey, M. G., Ouellette, M. E., . . . Baures, P. W. (2019). An Inhibitor of Fatty Acid Synthase Thioesterase Domain with Improved Cytotoxicity against Breast Cancer Cells and Stability in Plasma. *J Pharmacol Exp Ther*, *371*(1), 171-185. doi:10.1124/jpet.119.258947

Ma, M., Shi, X., & Lin, S. (2020). Heterologous expression and cell membrane localization of dinoflagellate opsins (rhodopsin proteins) in mammalian cells. *Marine Life Science & Technology*, *2*(3), 302-308. doi:10.1007/s42995-020-00043-1

Macpherson, G. R., Burton, I. W., LeBlanc, P., Walter, J. A., & Wright, J. L. (2003). Studies of the biosynthesis of DTX-5a and DTX-5b by the dinoflagellate Prorocentrum maculosum: regiospecificity of the putative Baeyer-Villigerase and insertion of a single amino acid in a polyketide chain. *J Org Chem*, *68*(5), 1659-1664. doi:10.1021/jo0204754

Mansour, M. P., Volkman, J. K., Jackson, A. E., & Blackburn, S. I. (1999). THE FATTY ACID AND STEROL COMPOSITION OF FIVE MARINE DINOFLAGELLATES. *Journal of Phycology*, *35*(4), 710-720. doi:10.1046/j.1529-8817.1999.3540710.x

Marechal, E., Block, M. A., Dorne, A.-J., Douce, R., & Joyard, J. (1997). Lipid synthesis and metabolism in the plastid envelope. *Physiologia Plantarum*, *100*(1), 65-77. doi:10.1111/j.1399-3054.1997.tb03455.x

Marinov, G. K., & Lynch, M. (2015). Diversity and Divergence of Dinoflagellate Histone Proteins. *G3 (Bethesda)*, *6*(2), 397-422. doi:10.1534/g3.115.023275

Markovic, P., Roenneberg, T., & Morse, D. (1996). Phased protein synthesis at several circadian times does not change protein levels in Gonyaulax. *J Biol Rhythms*, *11*(1), 57-67. doi:10.1177/074873049601100106

Maselli, M., Anestis, K., Klemm, K., Hansen, P. J., & John, U. (2021). Retention of Prey Genetic Material by the Kleptoplastidic Ciliate Strombidium cf. basimorphum. *Front Microbiol*, *12*, 694508. doi:10.3389/fmicb.2021.694508

Mazumdar, J., & Striepen, B. (2007). Make it or take it: fatty acid metabolism of apicomplexan parasites. *Eukaryot Cell*, *6*(10), 1727-1735. doi:10.1128/EC.00255-07

McDaniel, R., Thamchaipenet, A., Gustafsson, C., Fu, H., Betlach, M., & Ashley, G. (1999). Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel “unnatural” natural products. *Proc Natl Acad Sci U S A*, *96*(5), 1846-1851. doi:10.1073/pnas.96.5.1846

Meng, Y., Van Wagoner, R. M., Misner, I., Tomas, C., & Wright, J. L. (2010). Structure and biosynthesis of amphidinol 17, a hemolytic compound from Amphidinium carterae. *J Nat Prod*, *73*(3), 409-415. doi:10.1021/np900616q

Meyer, J. M., Rödelsperger, C., Eichholz, K., Tillmann, U., Cembella, A., McGaughran, A., & John, U. (2015). Transcriptomic characterisation and genomic glimps into the toxigenic dinoflagellate Azadinium spinosum, with emphasis on polykeitde synthase genes. *BMC Genomics*, *16*, 27. doi:10.1186/s12864-014-1205-6

Miller, J. J., Delwiche, C. F., & Coats, D. W. (2012). Ultrastructure of Amoebophrya sp. and its changes during the course of infection. *Protist*, *163*(5), 720-745. doi:10.1016/j.protis.2011.11.007

Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A., & Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res*, *41*(12), e121. doi:10.1093/nar/gkt263

Mittra, B., Zamudio, J. R., Bujnicki, J. M., Stepinski, J., Darzynkiewicz, E., Campbell, D. A., & Sturm, N. R. (2008). The TbMTr1 spliced leader RNA cap 1 2'-O-ribose methyltransferase from Trypanosoma brucei acts with substrate specificity. *J Biol Chem*, *283*(6), 3161-3172. doi:10.1074/jbc.M707367200

Monroe, E. A., Johnson, J. G., Wang, Z., Pierce, R. K., & Van Dolah, F. M. (2010). CHARACTERIZATION AND EXPRESSION OF NUCLEAR-ENCODED POLYKETIDE SYNTHASES IN THE BREVETOXIN-PRODUCING DINOFLAGELLATE KARENIA BREVIS1. *Journal of Phycology*, *46*(3), 541-552. doi:10.1111/j.1529-8817.2010.00837.x

Montagnes, D. J. S., Lowe, C. D., Roberts, E. C., Breckels, M. N., Boakes, D. E., Davidson, K., . . . Watts, P. C. (2011). An introduction to the special issue: Oxyrrhis marina, a model organism. *Journal of Plankton Research*, *33*(4), 549-554. doi:10.1093/plankt/fbq121

Moore, B. S., & Hertweck, C. (2002). Biosynthesis and attachment of novel bacterial polyketide synthase starter units. *Nat Prod Rep*, *19*(1), 70-99. doi:10.1039/b003939j

Moore, R. B., Oborník, M., Janouskovec, J., Chrudimský, T., Vancová, M., Green, D. H., . . . Carter, D. A. (2008). A photosynthetic alveolate closely related to apicomplexan parasites. *Nature*, *451*(7181), 959-963. doi:10.1038/nature06635

Moreno Díaz de la Espina, S., Alverca, E., Cuadrado, A., & Franca, S. (2005). Organization of the genome and gene expression in a nuclear environment lacking histones and nucleosomes: the amazing dinoflagellates. *Eur J Cell Biol*, *84*(2-3), 137-149. Retrieved from PubMed

Morey, J. S., & Van Dolah, F. M. (2013). Global analysis of mRNA half-lives and de novo transcription in a dinoflagellate, Karenia brevis. *PloS one*, *8*(6), e66347. Retrieved from Google Scholar

Morse, D., Milos, P. M., Roux, E., & Hastings, J. W. (1989). Circadian regulation of bioluminescence in Gonyaulax involves translational control. *Proc Natl Acad Sci U S A*, *86*(1), 172-176.

Morse, D., & Mittag, M. (2000). Dinoflagellate luciferin-binding protein. *Methods Enzymol*, *305*, 258-276. doi:10.1016/s0076-6879(00)05493-8

Mosavi, L. K., Cammett, T. J., Desrosiers, D. C., & Peng, Z. Y. (2004). The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci*, *13*(6), 1435-1448. doi:10.1110/ps.03554604

Murray, S. A., Diwan, R., Orr, R. J., Kohli, G. S., & John, U. (2015). Gene duplication, loss and selection in the evolution of saxitoxin biosynthesis in alveolates. *Mol Phylogenet Evol*, *92*, 165-180. doi:10.1016/j.ympev.2015.06.017

Murugan, E., Kong, R., Sun, H., Rao, F., & Liang, Z. X. (2010). Expression, purification and characterization of the acyl carrier protein phosphodiesterase from Pseudomonas Aeruginosa. *Protein Expr Purif*, *71*(2), 132-138. doi:10.1016/j.pep.2010.01.007

Naik, R., Obiang-Obounou, B. W., Kim, M., Choi, Y., Lee, H. S., & Lee, K. (2014). Therapeutic strategies for metabolic diseases: Small-molecule diacylglycerol acyltransferase (DGAT) inhibitors. *ChemMedChem*, *9*(11), 2410-2424. doi:10.1002/cmdc.201402069

Nilsson, A. K., Johansson, O. N., Fahlberg, P., Kommuri, M., Töpel, M., Bodin, L. J., . . . Andersson, M. X. (2015). Acylated monogalactosyl diacylglycerol: prevalence in the plant kingdom and identification of an enzyme catalyzing galactolipid head group acylation in Arabidopsis thaliana. *Plant J*, *84*(6), 1152-1166. doi:10.1111/tpj.13072

Nosenko, T., Lidie, K. L., Van Dolah, F. M., Lindquist, E., Cheng, J.-F., & Bhattacharya, D. (2006). Chimeric plastid proteome in the Florida "red tide" dinoflagellate Karenia brevis. *Mol Biol Evol*, *23*(11), 2026-2038. doi:10.1093/molbev/msl074

Novoa, E. M., & de Pouplana, L. R. (2012). Speeding with control: codon usage, tRNAs, and ribosomes. *Trends in Genetics*, *28*(11), 574-581. Retrieved from Google Scholar

O'Brien, J., Hayder, H., Zayed, Y., & Peng, C. (2018). Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Front Endocrinol (Lausanne)*, *9*, 402. doi:10.3389/fendo.2018.00402

Oakley, B. R., & Dodge, J. D. (1976). Mitosis and cytokinesis in the dinoflagellate Amphidinium carterae. *Cytobios*, *17*(65), 35-46. Retrieved from https://pubmed.ncbi.nlm.nih.gov/1036286

Oborník, M., Modrý, D., Lukeš, M., Cernotíková-Stříbrná, E., Cihlář, J., Tesařová, M., . . . Lukeš, J. (2012). Morphology, ultrastructure and life cycle of Vitrella brassicaformis n. sp., n. gen., a novel chromerid from the Great Barrier Reef. *Protist*, *163*(2), 306-323. doi:10.1016/j.protis.2011.09.001

Ohlrogge, J. B., & Jaworski, J. G. (1997). REGULATION OF FATTY ACID SYNTHESIS. *Annu Rev Plant Physiol Plant Mol Biol*, *48*, 109-136. doi:10.1146/annurev.arplant.48.1.109

Owen, J. G., Copp, J. N., & Ackerley, D. F. (2011). Rapid and flexible biochemical assays for evaluating 4'-phosphopantetheinyl transferase activity. *Biochem J*, *436*(3), 709-717. doi:10.1042/BJ20110321

Patterson, D. J. (1999). The Diversity of Eukaryotes. *Am Nat*, *154*(S4), S96-S124. doi:10.1086/303287

Paz, B., Daranas, A. H., Norte, M., Riobó, P., Franco, J. M., & Fernández, J. J. (2008). Yessotoxins, a group of marine polyether toxins: an overview. *Mar Drugs*, *6*(2), 73-102. doi:10.3390/md20080005

Peden, J. F. (2000). *Analysis of codon usage.* Citeseer, University of Nottingham.

Pelletreau, K. N., Bhattacharya, D., Price, D. C., Worful, J. M., Moustafa, A., & Rumpho, M. E. (2011). Sea slug kleptoplasty and plastid maintenance in a metazoan. *Plant Physiol*, *155*(4), 1561-1565. doi:10.1104/pp.111.174078

Peltomaa, E., Hällfors, H., & Taipale, S. J. (2019). Comparison of Diatoms and Dinoflagellates from Different Habitats as Sources of PUFAs. *Mar Drugs*, *17*(4), E233. doi:10.3390/md17040233

Peng, J., Place, A. R., Yoshida, W., Anklin, C., & Hamann, M. T. (2010). Structure and absolute configuration of karlotoxin-2, an ichthyotoxin from the marine dinoflagellate Karlodinium veneficum. *J Am Chem Soc*, *132*(10), 3277-3279. doi:10.1021/ja9091853

Piel, J. (2002). A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles. *Proc Natl Acad Sci U S A*, *99*(22), 14002-14007. doi:10.1073/pnas.222481399

Pincemin, J. M., & Gayol, P. (1978). Naked and Thecate Swarmers in Clonal Pyrocystis fusiformis:Systematic and Physiology Problems. *Archiv für Protistenkunde*, *120*(4), 401-408. doi:10.1016/s0003-9365(78)80031-1

Place, A. R., Bai, X., Kim, S., Sengco, M. R., & Wayne Coats, D. (2009). DINOFLAGELLATE HOST-PARASITE STEROL PROFILES DICTATE KARLOTOXIN SENSITIVITY(1). *J Phycol*, *45*(2), 375-385. doi:10.1111/j.1529-8817.2009.00649.x

Potapov, V., Fu, X., Dai, N., Corrêa, I. R., Tanner, N. A., & Ong, J. L. (2018). Base modifications affecting RNA polymerase and reverse transcriptase fidelity. *Nucleic Acids Res*, *46*(11), 5753-5763. doi:10.1093/nar/gky341

Probert, I., Siano, R., Poirier, C., Decelle, J., Biard, T., Tuji, A., . . . Not, F. (2014). Brandtodinium gen. nov. and B. nutricula comb. Nov. (Dinophyceae), a dinoflagellate commonly found in symbiosis with polycystine radiolarians. *J Phycol*, *50*(2), 388-399. doi:10.1111/jpy.12174

Rae, P. M. (1973). 5-Hydroxymethyluracil in the DNA of a dinoflagellate. *Proc Natl Acad Sci U S A*, *70*(4), 1141-1145.

Raghukumar, S. (2002). Ecology of the marine protists, the Labyrinthulomycetes (Thraustochytrids and Labyrinthulids). *European Journal of Protistology*, *38*(2), 127-145. doi:10.1078/0932-4739-00832

Rasmussen, S. A., Binzer, S. B., Hoeck, C., Meier, S., de Medeiros, L. S., Andersen, N. G., . . . Larsen, T. O. (2017). Karmitoxin: An Amine-Containing Polyhydroxy-Polyene Toxin from the Marine Dinoflagellate Karlodinium armiger. *J Nat Prod*, *80*(5), 1287-1293. doi:10.1021/acs.jnatprod.6b00860

Rausch, C., Hoof, I., Weber, T., Wohlleben, W., & Huson, D. H. (2007). Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol Biol*, *7*, 78. doi:10.1186/1471-2148-7-78

Regaudie-de-Gioux, A., Lasternas, S. Ã., AgustÃ-, S., & Duarte, C. M. (2014). Comparing marine primary production estimates through different methods and

development of conversion equations. *Frontiers in Marine Science*, *1*. doi:10.3389/fmars.2014.00019
Remize, M., Planchon, F., Garnier, M., Loh, A. N., Le Grand, F., Bideau, A., . . . Soudant, P. (2021). A $^{13}CO_2$ Enrichment Experiment to Study the Synthesis Pathways of Polyunsaturated Fatty Acids of the Haptophyte Tisochrysis lutea. *Mar Drugs*, *20*(1), 22. doi:10.3390/md20010022
Ren, L.-j., Chen, S.-l., Geng, L.-j., Ji, X.-j., Xu, X., Song, P., . . . Huang, H. (2018). Exploring the function of acyltransferase and domain replacement in order to change the polyunsaturated fatty acid profile of Schizochytrium sp. *Algal Research*, *29*, 193-201. doi:10.1016/j.algal.2017.11.021
Rodriguez, J. D., Haq, S., Bachvaroff, T., Nowak, K. F., Nowak, S. J., Morgan, D., . . . Smith, S. M. (2017). Identification of a vacuolar proton channel that triggers the bioluminescent flash in dinoflagellates. *PLoS One*, *12*(2), e0171594. doi:10.1371/journal.pone.0171594
Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol*, *5*, 172. doi:10.3389/fmicb.2014.00172
Roy, S., Jagus, R., & Morse, D. (2018). Translation and Translational Control in Dinoflagellates. *Microorganisms*, *6*(2), E30. doi:10.3390/microorganisms6020030
Roy, S., & Morse, D. (2013). Transcription and Maturation of mRNA in Dinoflagellates. *Microorganisms*, *1*, 71-99. doi:10.3390/microorganisms1010071
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Retrieved from http://books.google.com/books?id=-mPwAAAAMAAJ&hl=&source=gbs_api
Sasaki, M., Matsumori, N., Maruyama, T., Nonomura, T., Murata, M., Tachibana, K., & Yasumoto, T. (1996). The Complete Structure of Maitotoxin, Part I: Configuration of the C1C14 Side Chain. *Angewandte Chemie International Edition in English*, *35*(15), 1672-1675. doi:10.1002/anie.199616721
Satake, M., Morohashi, A., Oguri, H., Oishi, T., Hirama, M., Harada, N., & Yasumoto, T. (1997). The absolute configuration of ciguatoxin. *Journal of the American Chemical Society*, *119*(46), 11325-11326. Retrieved from https://pubs.acs.org/doi/full/10.1021/ja972482t
Sato, N. (2020). Complex origins of chloroplast membranes with photosynthetic machineries: multiple transfers of genes from divergent organisms at different times or a single endosymbiotic event. *J Plant Res*, *133*(1), 15-33. doi:10.1007/s10265-019-01157-z
Sato, S. (2011). The apicomplexan plastid and its evolution. *Cell Mol Life Sci*, *68*(8), 1285-1296. doi:10.1007/s00018-011-0646-1
Schnepf, E., & ElbräChter, M. (1999). Dinophyte chloroplasts and phylogeny - A review. *Grana*, *38*(2-3), 81-97. doi:10.1080/00173139908559217
Schröder-Lorenz, A., & Rensing, L. (1987). Circadian changes in protein-synthesis rate and protein phosphorylation in cell-free extracts of Gonyaulax polyedra. *Planta*, *170*(1), 7-13. doi:10.1007/BF00392374

Schümann, J., & Hertweck, C. (2006). Advances in cloning, functional analysis and heterologous expression of fungal polyketide synthase genes. *J Biotechnol*, *124*(4), 690-703. doi:10.1016/j.jbiotec.2006.03.046

Seki, T., Satake, M., Mackenzie, L., Kaspar, H. F., & Yasumoto, T. (1995). Gymnodimine, a new marine toxin of unprecedented structure isolated from New Zealand oysters and the dinoflagellate, Gymnodinium sp. *Tetrahedron letters*, *36*(39), 7093-7096. Retrieved from Google Scholar

Sheng, J., Malkiel, E., Katz, J., Adolf, J. E., & Place, A. R. (2010). A dinoflagellate exploits toxins to immobilize prey prior to ingestion. *Proc Natl Acad Sci U S A*, *107*(5), 2082-2087. doi:10.1073/pnas.0912254107

Shi, X., Lin, X., Li, L., Li, M., Palenik, B., & Lin, S. (2017). Transcriptomic and microRNAomic profiling reveals multi-faceted mechanisms to cope with phosphate stress in a dinoflagellate. *ISME J*, *11*(10), 2209-2218. doi:10.1038/ismej.2017.81

Shoguchi, E., Shinzato, C., Kawashima, T., Gyoja, F., Mungpakdee, S., Koyanagi, R., . . . Satoh, N. (2013). Draft assembly of the Symbiodinium minutum nuclear genome reveals dinoflagellate gene structure. *Curr Biol*, *23*(15), 1399-1408. doi:10.1016/j.cub.2013.05.062

Shuman, S. (2002). What messenger RNA capping tells us about eukaryotic evolution. *Nat Rev Mol Cell Biol*, *3*(8), 619-625. doi:10.1038/nrm880

Sieber, S. A., & Marahiel, M. A. (2005). Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem Rev*, *105*(2), 715-738. doi:10.1021/cr0301191

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210-3212. doi:10.1093/bioinformatics/btv351

Skovgaard, A., Meneses, I., & Angélico, M. M. (2009). Identifying the lethal fish egg parasite Ichthyodinium chabelardi as a member of Marine Alveolate Group I. *Environ Microbiol*, *11*(8), 2030-2041. doi:10.1111/j.1462-2920.2009.01924.x

Snyder, R. V., Gibbs, P. D. L., Palacios, A., Abiy, L., Dickey, R., Lopez, J. V., & Rein, K. S. (2003). Polyketide synthase genes from marine dinoflagellates. *Mar Biotechnol (NY)*, *5*(1), 1-12. doi:10.1007/s10126-002-0077-y

Song, B., Morse, D., Song, Y., Fu, Y., Lin, X., Wang, W., . . . Lin, S. (2017). Comparative Genomics Reveals Two Major Bouts of Gene Retroposition Coinciding with Crucial Periods of Symbiodinium Evolution. *Genome Biol Evol*, *9*(8), 2037-2047. doi:10.1093/gbe/evx144

Sonnenschein, E. C., Pu, Y., Beld, J., & Burkart, M. D. (2016). Phosphopantetheinylation in the green microalgae Chlamydomonas reinhardtii. *Journal of Applied Phycology*, *28*(6), 3259-3267. doi:10.1007/s10811-016-0875-7

Sørensen, H. P., & Mortensen, K. K. (2005). Advanced genetic strategies for recombinant protein expression in Escherichia coli. *J Biotechnol*, *115*(2), 113-128. doi:10.1016/j.jbiotec.2004.08.004

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312-1313. doi:10.1093/bioinformatics/btu033

Stephens, T. G., González-Pech, R. A., Cheng, Y., Mohamed, A. R., Burt, D. W., Bhattacharya, D., . . . Chan, C. X. (2020). Genomes of the dinoflagellate Polarella glacialis encode tandemly repeated single-exon genes with adaptive functions. *BMC Biol*, *18*(1), 56. doi:10.1186/s12915-020-00782-8

Strassert, J. F. H., Karnkowska, A., Hehenberger, E., Del Campo, J., Kolisko, M., Okamoto, N., . . . Keeling, P. J. (2018). Single cell genomics of uncultured marine alveolates shows paraphyly of basal dinoflagellates. *ISME J*, *12*(1), 304-308. doi:10.1038/ismej.2017.167

Studier, F. W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif*, *41*(1), 207-234. doi:10.1016/j.pep.2005.01.016

Sun, S., Chen, J., Li, W., Altintas, I., Lin, A., Peltier, S., . . . Wooley, J. (2011). Community cyberinfrastructure for Advanced Microbial Ecology Research and Analysis: the CAMERA resource. *Nucleic Acids Res*, *39*(Database issue), D546-51. doi:10.1093/nar/gkq1102

Takahashi, H., Kumagai, T., Kitani, K., Mori, M., Matoba, Y., & Sugiyama, M. (2007). Cloning and characterization of a Streptomyces single module type non-ribosomal peptide synthetase catalyzing a blue pigment synthesis. *J Biol Chem*, *282*(12), 9073-9081. doi:10.1074/jbc.M611319200

Tatsuta, T., Scharwey, M., & Langer, T. (2014). Mitochondrial lipid trafficking. *Trends Cell Biol*, *24*(1), 44-52. doi:10.1016/j.tcb.2013.07.011

Taylor, F. J. R., Hoppenrath, M., & Saldarriaga, J. F. (2008). Dinoflagellate diversity and distribution. *Biodiversity and Conservation*, *17*(2), 407-418. doi:10.1007/s10531-007-9258-3

Tengs, T., Dahlberg, O. J., Shalchian-Tabrizi, K., Klaveness, D., Rudi, K., Delwiche, C. F., & Jakobsen, K. S. (2000). Phylogenetic analyses indicate that the 19'Hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. *Mol Biol Evol*, *17*(5), 718-729. Retrieved from PubMed

Tillett, D., Dittmann, E., Erhard, M., von Döhren, H., Börner, T., & Neilan, B. A. (2000). Structural organization of microcystin biosynthesis in Microcystis aeruginosa PCC7806: an integrated peptide-polyketide synthetase system. *Chem Biol*, *7*(10), 753-764. doi:10.1016/s1074-5521(00)00021-1

Tippit, D. H., & Pickett-Heaps, J. D. (1976). Apparent amitosis in the binucleate dinoflagellate Peridinium balticum. *J Cell Sci*, *21*(2), 273-289. doi:10.1242/jcs.21.2.273

Tong, L. (2005). Acetyl-coenzyme A carboxylase: crucial metabolic enzyme and attractive target for drug discovery. *Cellular and Molecular Life Sciences*, *62*(16), 1784-1803. doi:10.1007/s00018-005-5121-4

Twiner, M. J., Flewelling, L. J., Fire, S. E., Bowen-Stevens, S. R., Gaydos, J. K., Johnson, C. K., . . . Rowles, T. K. (2012). Comparative analysis of three brevetoxin-associated bottlenose dolphin (Tursiops truncatus) mortality events in the Florida Panhandle region (USA). *PLoS One*, *7*(8), e42974. doi:10.1371/journal.pone.0042974

Valastyan, J. S., Tota, M. R., Taylor, I. R., Stergioula, V., Hone, G. A. B., Smith, C. D., . . . Bassler, B. L. (2020). Discovery of PqsE Thioesterase Inhibitors for Pseudomonas aeruginosa Using DNA-Encoded Small Molecule Library Screening. *ACS Chem Biol*, *15*(2), 446-456. doi:10.1021/acschembio.9b00905

Van Dolah, F. M., Kohli, G. S., Morey, J. S., & Murray, S. A. (2017). Both modular and single-domain Type I polyketide synthases are expressed in the brevetoxin-producing dinoflagellate, Karenia brevis (Dinophyceae). *J Phycol*, *53*(6), 1325-1339. doi:10.1111/jpy.12586

Van Dolah, F. M., Zippay, M. L., Pezzolesi, L., Rein, K. S., Johnson, J. G., Morey, J. S., . . . Pistocchi, R. (2013). Subcellular localization of dinoflagellate polyketide synthases and fatty acid synthase activity. *J Phycol*, *49*(6), 1118-1127. doi:10.1111/jpy.12120

Van Dolah, F. M., Morey, J. S., Milne, S., Ung, A., Anderson, P. E., & Chinain, M. (2020). Transcriptomic analysis of polyketide synthases in a highly ciguatoxic dinoflagellate, Gambierdiscus polynesiensis and low toxicity Gambierdiscus pacificus, from French Polynesia. *PLOS ONE*, *15*(4), e0231400. Retrieved from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0231400

Van Wagoner, R. M., Deeds, J. R., Satake, M., Ribeiro, A. A., Place, A. R., & Wright, J. L. (2008). Isolation and characterization of karlotoxin 1, a new amphipathic toxin from Karlodinium veneficum. *Tetrahedron Lett*, *49*(45), 6457-6461. doi:10.1016/j.tetlet.2008.08.103

Van Wagoner, R. M., Deeds, J. R., Tatters, A. O., Place, A. R., Tomas, C. R., & Wright, J. L. (2010). Structure and relative potency of several karlotoxins from Karlodinium veneficum. *J Nat Prod*, *73*(8), 1360-1365. doi:10.1021/np100158r

Van Wagoner, R. M., Satake, M., & Wright, J. L. (2014). Polyketide biosynthesis in dinoflagellates: what makes it different. *Nat Prod Rep*, *31*(9), 1101-1137. doi:10.1039/c4np00016a

Veldhuis, M. J. W., & Kraay, G. W. (2000). Application of flow cytometry in marine phytoplankton research: current applications and future perspectives. *Scientia Marina*, *64*(2), 121-134. doi:10.3989/scimar.2000.64n2121

Verma, A., Barua, A., Ruvindy, R., Savela, H., Ajani, P. A., & Murray, S. A. (2019). The Genetic Basis of Toxin Biosynthesis in Dinoflagellates. *Microorganisms*, *7*(8), E222. doi:10.3390/microorganisms7080222

Walsh, C. J., Butawan, M., Yordy, J., Ball, R., Flewelling, L., de Wit, M., & Bonde, R. K. (2015). Sublethal red tide toxin exposure in free-ranging manatees (Trichechus manatus) affects the immune system through reduced lymphocyte proliferation responses, inflammation, and oxidative stress. *Aquat Toxicol*, *161*, 73-84. doi:10.1016/j.aquatox.2015.01.019

Wang, D. Z. (2008). Neurotoxins from marine dinoflagellates: a brief review. *Mar Drugs*, *6*(2), 349-371. doi:10.3390/md20080016

Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L., & Sivonen, K. (2014). Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proc Natl Acad Sci U S A*, *111*(25), 9259-9264. doi:10.1073/pnas.1401734111

Wang, S., Lan, C., Wang, Z., Wan, W., Cui, Q., & Song, X. (2020). PUFA-synthase-specific PPTase enhanced the polyunsaturated fatty acid biosynthesis via the

polyketide synthase pathway in Aurantiochytrium. *Biotechnol Biofuels*, *13*, 152. doi:10.1186/s13068-020-01793-x
Wang, S. P., Deng, L., Ho, C. K., & Shuman, S. (1997). Phylogeny of mRNA capping enzymes. *Proceedings of the National Academy of Sciences*, *94*(18), 9573-9578. Retrieved from Google Scholar
Wang, X., Fosse, H. K., Li, K., Chauton, M. S., Vadstein, O., & Reitan, K. I. (2019). Influence of Nitrogen Limitation on Lipid Accumulation and EPA and DHA Content in Four Marine Microalgae for Possible Use in Aquafeed. *Frontiers in Marine Science*, *6*. doi:10.3389/fmars.2019.00095
Warns, A., Hense, I., & Kremp, A. (2013). Modelling the life cycle of dinoflagellates: a case study with Biecheleria baltica. *Journal of Plankton Research*, *35*(2), 379-392. doi:10.1093/plankt/fbs095
William, R. E. (1962). Dinoflagellates and Their Use in Petroleum Geology: ABSTRACT. *AAPG Bulletin*, *46*. doi:10.1306/bc7437a3-16be-11d7-8645000102c1865d
Williams, E., Place, A., & Bachvaroff, T. (2017). Transcriptome Analysis of Core Dinoflagellates Reveals a Universal Bias towards "GC" Rich Codons. *Mar Drugs*, *15*(5). doi:10.3390/md15050125
Williams, E. P., Bachvaroff, T. R., & Place, A. R. (2021). A Global Approach to Estimating the Abundance and Duplication of Polyketide Synthase Domains in Dinoflagellates. *Evol Bioinform Online*, *17*, 11769343211031871. doi:10.1177/11769343211031871
Williams, E. P., & Place, A. R. (2014). *Proliferation of 5-hydroxymethyl uracil in the genomes of dinoflagellates is synapomorphic to dinokaryon containing species*. Proceedings from Proceedings of the 15th International Conference on Harmful Algae, Changwon, Korea, Busan, Korea.
Williams, E. P., Bachvaroff, T. R., & Place, A. R. (2020). *The Phosphopantetheinyl Transferases in Dinoflagellates*. Proceedings from THE 18TH INTERNATIONAL CONFERENCE ON HARMFUL ALGAE, Nantes.
Wisecaver, J. H., Brosnahan, M. L., & Hackett, J. D. (2013). Horizontal gene transfer is a significant driver of gene innovation in dinoflagellates. *Genome Biol Evol*, *5*(12), 2368-2381. doi:10.1093/gbe/evt179
Withers, N. W. (1982). Ciguatera fish poisoning. *Annu Rev Med*, *33*, 97-111. doi:10.1146/annurev.me.33.020182.000525
Wright, J. L. C., Hu, T., McLachlan, J. L., Needham, J., & Walter, J. A. (1996). Biosynthesis of DTX-4: confirmation of a polyketide pathway, proof of a baeyer− villiger oxidation step, and evidence for an unusual carbon deletion process. *Journal of the American Chemical Society*, *118*(36), 8757-8758. Retrieved from https://pubs.acs.org/doi/full/10.1021/ja961715y
Yamada, N., Bolton, J. J., Trobajo, R., Mann, D. G., Dąbek, P., Witkowski, A., . . . Kroth, P. G. (2019). Discovery of a kleptoplastic 'dinotom' dinoflagellate and the unique nuclear dynamics of converting kleptoplastids to permanent plastids. *Sci Rep*, *9*(1), 10474. doi:10.1038/s41598-019-46852-y
Yamada, N., Sym, S. D., & Horiguchi, T. (2017). Identification of Highly Divergent Diatom-Derived Chloroplasts in Dinoflagellates, Including a Description of

Durinskia kwazulunatalensis sp. nov. (Peridiniales, Dinophyceae). *Mol Biol Evol*, *34*(6), 1335-1351. doi:10.1093/molbev/msx054
Yan, T. H. K., Wu, Z., Kwok, A. C. M., & Wong, J. T. Y. (2020). Knockdown of Dinoflagellate Condensin CcSMC4 Subunit Leads to S-Phase Impediment and Decompaction of Liquid Crystalline Chromosomes. *Microorganisms*, *8*(4), E565. doi:10.3390/microorganisms8040565
Yazawa, K. (1996). Production of eicosapentaenoic acid from marine bacteria. *Lipids*, *31 Suppl*, S297-300. doi:10.1007/BF02637095
Yi, Z., Xu, M., Di, X., Brynjolfsson, S., & Fu, W. (2017). Exploring Valuable Lipids in Diatoms. *Frontiers in Marine Science*, *4*. doi:10.3389/fmars.2017.00017
Yih, W., & Coats, D. W. (2000). Infection of Gymnodinium sanguineum by the dinoflagellate Amoebophrya sp.: effect of nutrient environment on parasite generation time, reproduction, and infectivity. *J Eukaryot Microbiol*, *47*(5), 504-510. doi:10.1111/j.1550-7408.2000.tb00082.x
Yoon, H. S., Hackett, J. D., & Bhattacharya, D. (2002). A single origin of the peridinin- and fucoxanthin-containing plastids in dinoflagellates through tertiary endosymbiosis. *Proc Natl Acad Sci U S A*, *99*(18), 11724-11729. doi:10.1073/pnas.172234799
Zeytuni, N., & Zarivach, R. (2012). Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. *Structure*, *20*(3), 397-405. doi:10.1016/j.str.2012.01.006
Zhang, H., Hou, Y., Miranda, L., Campbell, D. A., Sturm, N. R., Gaasterland, T., & Lin, S. (2007). Spliced leader RNA trans-splicing in dinoflagellates. *Proc Natl Acad Sci U S A*, *104*(11), 4618-4623. doi:10.1073/pnas.0700258104
Zhang, H., Zhuang, Y., Gill, J., & Lin, S. (2013). Proof that dinoflagellate spliced leader (DinoSL) is a useful hook for fishing dinoflagellate transcripts from mixed microbial samples: Symbiodinium kawagutii as a case study. *Protist*, *164*(4), 510-527. doi:10.1016/j.protis.2013.04.002
Zhang, H., Campbell, D. A., Sturm, N. R., & Lin, S. (2009). Dinoflagellate spliced leader RNA genes display a variety of sequences and genomic arrangements. *Mol Biol Evol*, *26*(8), 1757-1771. doi:10.1093/molbev/msp083
Zhang, Z., Green, B. R., & Cavalier-Smith, T. (1999). Single gene circles in dinoflagellate chloroplast genomes. *Nature*, *400*(6740), 155-159. doi:10.1038/22099
Zhao, L. S., Li, K., Wang, Q. M., Song, X. Y., Su, H. N., Xie, B. B., . . . Zhang, Y. Z. (2017). Nitrogen Starvation Impacts the Photosynthetic Performance of Porphyridium cruentum as Revealed by Chlorophyll a Fluorescence. *Sci Rep*, *7*(1), 8542. doi:10.1038/s41598-017-08428-6
Zhu, G., Li, Y., Cai, X., Millership, J. J., Marchewka, M. J., & Keithly, J. S. (2004). Expression and functional characterization of a giant Type I fatty acid synthase (CpFAS1) gene from Cryptosporidium parvum. *Mol Biochem Parasitol*, *134*(1), 127-135. doi:10.1016/j.molbiopara.2003.11.011
Zykova, T. Y., Levitsky, V. G., Belyaeva, E. S., & Zhimulev, I. F. (2018). Polytene Chromosomes - A Portrait of Functional Organization of the Drosophila Genome. *Curr Genomics*, *19*(3), 179-191. doi:10.2174/1389202918666171016123830