ABSTRACT

Title of Dissertation:	ESTABLISHING A MODEL: *AMPHIDINIUM CARTERAE* AS A TEMPLATE FOR DINOFLAGELLATE GENOMICS & EXPRESSION
	Miranda Judd, Doctor of Philosophy, 2025
Dissertation directed by:	Dr. Allen R. Place, Department of Marine Estuarine and Environmental Sciences Dr. Tsvetan Bachvaroff, Department of Marine Estuarine and Environmental Sciences

Dinoflagellates are globally important protists that underpin marine food webs, drive coral symbioses, and generate harmful algal blooms, yet their unique genomic architecture has hindered their adoption as tractable model systems. This dissertation advances *Amphidinium carterae* as a reference dinoflagellate by combining ecological, molecular, and genomic approaches to address key barriers in the field.

First, the long-term laboratory microbiome of *A. carterae* was characterized, assembling 15 complete bacterial genomes with long-read sequencing. These analyses revealed antibiotic-resistant taxa, metabolic pathways linked to nutrient cycling, and potential interactions with host physiology, underscoring the importance of defining the microbiome when developing experimental models.

Next, a functional genomics tool was established by delivering antisense morpholinos targeting the translation initiation factor eIF4E-1a. This knockdown system reduced protein abundance without severely impacting viability, providing proof-of-principle for genetic manipulation in dinoflagellates.

The dinoflagellate approach to transcription was described using synchronized cultures, integrating Nanopore direct RNA sequencing, ribosome profiling, and translation assays over a diel cycle. Results showed that gene regulation is dominated by post-transcriptional mechanisms, including dynamic translational control and pervasive RNA modifications (m^6A, m^5C, and pseudouridine). Distinct functional enrichments were observed across Growth, Synthesis, and Division phases, highlighting translational orchestration as the primary driver of cellular rhythms.

Finally, a 1.25 Gb long-read genome assembly for *A. carterae* was produced with curated annotation of tandemly duplicated genes, nested gene structures, and extensive repetitive regions. Tandem gene arrays were found to strongly correlate with transcript abundance, suggesting a dosage-based regulatory strategy in lieu of conventional transcriptional control. Across different genome regions there were strong differences in expression of different transcripts. Comparative analyses confirmed these features as conserved across dinoflagellates, while also highlighting the prevalence of lineage-specific “dark genes” with unknown function.

Together, these studies expand genomic resources, functional tools, and expression analyses for *A. carterae*, establishing it as a tractable system for probing dinoflagellate biology. More broadly, they illuminate alternative strategies of genome

organization and regulation in eukaryotes, providing a foundation for future work on dinoflagellate ecology, evolution, and the mechanisms underlying their environmental impacts.

ESTABLISHING A MODEL: *AMPHIDINIUM CARTERAE* AS A TEMPLATE FOR DINOFLAGELLATE GENOMICS & EXPRESSION

by

Miranda Judd

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
2025

Advisory Committee:
Allen R. Place, Chair
Tsvetan Bachvaroff, Chair
Kathleen Cusick
Yantao Li
Helen Dooley

Dedication

This work is dedicated to my wonderful and supportive family—my parents, Leeann and Stephen Kelly-Judd; my siblings, Seth and Mariel; and my grandmother, Alice Judd—whose encouragement has meant so much to me. I am especially grateful to my husband, Brenden Caffey, for his unwavering love and support throughout my graduate studies. Finally, I would like to thank my dog, Ricky Bobby, for lifting my spirits during this journey, even if his constant whining while I wrote was less than helpful.

Acknowledgements

The hardware used in the computational studies is part of the UMBC High Performance Computing Facility (HPCF). The facility is supported by the U.S. National Science Foundation through the MRI program (grant nos. CNS–0821258, CNS–1228778, OAC–1726023, and CNS–1920079) and the SCREMS program (grant no. DMS–0821311), with additional substantial support from the University of Maryland, Baltimore County (UMBC). See hpcf.umbc.edu (accessed on 7 July 2024) for more information on HPCF and the projects using its resources. The BioAnalytical Services Laboratory (BASLab) at the Institute of Marine and Environmental Technology (IMET) was used for Nanopore MinION long-read sequencing. See https://www.umces.edu/baslab (accessed on 7 July 2024) for more information on the BASLab capabilities. Cold Spring Harbor Laboratory services were used for Nanopore PromethION long-read sequencing. See https://www.cshl.edu/research/cancer/sequencing-technologies-and-analysis/ for more information on Cold Spring Harbor Laboratory services.

Stipend funding included contributions from the UMCP Wylie Dissertation Fellowship, the Vice President for Education, the MEES program, the Florida Red Tide Mitigation and Technology Development Initiative, the IMET BioAnalytical Services Laboratory, and Allen R. Place.

Table of Contents

List of Tables

Supplemental Tables:

List of Figures

Supplemental Figure ii. Western blot analyses for eIF4E-1a concentrations within control and treated cells at 96 hours post-MO

Supplemental Figure iii. BioAnalyzer results for extracted and isolated RNA before poly-A selection and preparation for nanopore direct RNA sequencing

Supplemental Figure iv. Western blots of *A. carterae* cell lysates over the diel cycle probed with monoclonal mouse anti-puromycin antibody and a secondary goat anti-mouse HRP monoclonal antibody

Supplemental Figure v. Distribution of Ribo-Seq read mapping

Supplemental Figure vi. Tandem gene structure characteristics. Left: exon lengths, Center: intron lengths (log transformed), and Right: exon number

Supplemental Figure vii. D-Genies mapping of the present assembly to the Li *et al* 2024 assembly of another strain of *A. carterae*

Supplemental Figure viii. Map of tandemly repeated actin gene copies in the present assembly (top) and the Li *et al* 2024 assembly (bottom).

Supplemental Figure ix. Number of exons in single-copy genes and their associated FPKM

List of Abbreviations

HAB – Harmful Algal Bloom

MBNP – Major basic Nuclear Protein

DVNP – Dinoflagellate/Viral Nucleoprotein

DNA – Deoxyribonucleic Acid

RNA – Ribonucleic Acid

RNA-seq – RNA Sequencing

DEG – Differentially Expressed Gene

TF – Transcription Factor

PCR – Polymerase Chain Reaction

LogFC – Logarithmic Fold-Change

MO – morpholino

TPM – Transcripts per Million

FPKM – Fragments Per Kilobase Million

ESAW – Enriched Water, Artificial Water

m^6A – N^6-Methyladenosine

m^1A – N^1-Methyladenosine

m^5C – 5-methylcytosine

pseU – Pseudouridine

Chapter 1: Introduction

What are Dinoflagellates

Dinoflagellates (phylum Dinoflagellata or Dinophyceae) are a monophyletic clade within the Alveolata supergroup—alongside apicomplexans and ciliates (Bachvaroff *et al.* 2014; Hackett *et al.* 2004; Price & Bhattacharya 2017; Riding *et al.* 2022). Their evolutionary relationships are supported by molecular phylogenetics, which consistently confirm this grouping. Early diverging (basal) lineages include the monophyletic Syndiniales, which are linked to the core dinoflagellate clade. Within the core dinoflagellates the 'naked' or Gymnodinian group is paraphyletic to the thecate or 'armored' clade (Gómez 2020; Janouskovec *et al.* 2017; Spector 1984). The fossil record—especially dinoflagellate cysts—traces their ancestry back to the mid-Triassic, displaying earliest unambiguous fossil evidence of dinoflagellate cysts, even though some studies have concluded that molecular data suggest they may have originated earlier (Hoppenrath 2017; Janouskovec *et al.* 2017; Mertens *et al.* 2012).

Ecologically and taxonomically, dinoflagellates are incredibly versatile, comprising roughly 2,000 to 6,000 described species across up to 550 genera, depending on the classification used (Gómez 2020, 2012; Riding *et al.* 2022). Dinoflagellates thrive in marine, brackish, and freshwater environments, even colonizing sea ice and snow, with transitions between marine and freshwater habitats occurring repeatedly throughout their history. Functionally, they span a wide trophic spectrum—from strict autotrophs and mixotrophs to heterotrophs, parasites, and predators—with many species performing mixotrophy, i.e., photosynthesis supplemented by phagotrophy (Kim & Kim 2007).

Dinoflagellate photosynthesis is unique in that many species use peridinin – a hallmark carotenoid of the ancestral photosynthetic dinoflagellate lineage. Peridinin is the main accessory carotenoid in the peridinin–chlorophyll–protein (PCP) complexes that uniquely broaden light capture into the blue-green spectrum (≈470–550 nm) and channel energy efficiently to chlorophyll *a* (Bachvaroff *et al.* 2006; Dorrell & Howe 2015; Morden & Sherwood 2002; Takishita 2004; Yoon *et al.* 2002). This pigment is widespread among peridinin-containing dinoflagellates, underscoring its status as the primary plastid type. However, some groups—such as the genera *Karenia*, *Karlodinium*, and *Takayama*—have replaced their original peridinin plastid through tertiary endosymbiosis with a haptophyte-derived plastid, which uses the accessory carotenoid 19'-hexanoyloxyfucoxanthin (Tengs *et al.* 2000; Yoon *et al.* 2002). Interestingly, these "fucoxanthin dinoflagellates" retain nuclear-encoded machinery inherited from their peridinin ancestors, such as RNA editing and plastid transcript-processing genes, even after plastid replacement (Tengs *et al.* 2000). Several other non-canonical and temporary (kleptoplastid) systems also occur in other dinoflagellates species, which demonstrate multiple independent endosymbiosis or plastid replacement events, as well as intermediate stages (Burki *et al.* 2014; Kamikawa *et al.* 2015; Kim & Archibald 2010; Matsuo *et al.* 2022; Yamada *et al.* 2019). As a result, dinoflagellates today display a remarkable pigment diversity—from peridinin-dependent light harvesting in classic lineages to fucoxanthin-based systems in derived groups—illustrating both evolutionary flexibility and ecological adaptation in their plastid biology (Delwiche 1999).

Dinoflagellate ecological roles include serving as primary producers, prey and predators within planktonic food webs, and symbionts in coral reefs—most famously the Symbiodiniaceae (zooxanthellae) that power coral ecosystems. Mixotrophic dinoflagellates are versatile planktonic protists that combine photosynthesis and heterotrophic feeding—such as phagotrophy or kleptoplasty—to supplement nutritional uptake, often enabling them to thrive in fluctuating

environments (Adolf *et al.* 2006; Hansen 2011; Place *et al.* 2012). Some dinoflagellate species, like *Gambierdiscus* and *Ostreopsis*, are notorious for producing toxins leading to foodborne illnesses (e.g., ciguatera) or harmful algal blooms (HABs) affecting marine life and human health (Litaker *et al.* 2010; Patocka *et al.* 2015).

Some bloom-forming dinoflagellates produce a class of polyketide toxins called sterolysins. Sterolysins are potent membrane-active toxins produced by dinoflagellates such as *Karlodinium veneficum* and *Amphidinium carterae*, and their ecological significance is underscored by the foundational work of Allen R. Place. In a landmark study, Place *et al.* (2024) isolated two sterol-dependent toxins—abbotoxin and its chloro-derivative—from a long-term culture of *K. veneficum* and demonstrated that these karlotoxin congeners form large-conductance, cholesterol-binding membrane pores with pronounced ionic selectivity and voltage-dependent gating (Place *et al.* 2024). Earlier investigations showed that karlotoxins such as KmTx2 selectively bind to cholesterol and desmethyl sterols, lysing erythrocyte membranes rich in cholesterol while sparing those containing alternative sterols—an elegant mechanism for self-protection (Deeds & Place 2006). Research highlights sterolysins as cholesterol-dependent cytolysins, reinforcing their significance in marine predator–prey dynamics and environmental toxicology (Place *et al.* 2024, 2009)

Harmful algal blooms (HABs) caused by toxin-producing dinoflagellates—such as *Karenia brevis*, *Alexandrium* spp., and others—inflict substantial economic damage, particularly on coastal tourism. In the United States, HAB events cost the economy between $10–100 million annually, with individual large blooms resulting in tens of millions of dollars in lost revenue (Larkin & Adams 2007; National Centers for Coastal Ocean Science 2017). For example, the 2018 Florida red tide resulted in estimated losses of $2.7 billion to tourism-related businesses across the state, disrupting hotel occupancy, beach activities, and local hospitality services

(University of Florida Sea Grant 2024). In earlier decades, red tides on Florida's Gulf Coast in the 1970s caused tourism losses of $71 million in 1974 and $107 million in 1971, alongside significant fishing and cleanup costs (National Oceanic and Atmospheric Administration - National Marine Fisheries Service 2024). Even lower-concentration blooms can have outsized economic effects, due to the swift spread of public concern and cancellations—even when visual symptoms of blooms are mild. These recurring and growing losses underscore the urgent need for monitoring, forecasting, and mitigation strategies to safeguard public health and preserve coastal economies.

Thus, dinoflagellates not only present a fascinating fusion of evolutionary history, diverse morphology and feeding strategies, and profound ecological impact—from supporting coral reefs to driving both beneficial and harmful blooms across aquatic ecosystems, they also present a large problem to the bioeconomy. These characteristics emphasize why further research and understanding is necessary.

The Dinoflagellate Nucleus and Bizarre Cell Biology

Dinoflagellate genomes are estimated to be very large, with some measurements alluding to DNA content over hundreds of picograms per cell, and chromosome numbers ranging from a few to hundreds (Lin 2011). These measurements are typically measured via flow cytometry using common eukaryotic cells with known genome sizes as internal standards (Du *et al.* 2016; Stüken *et al.* 2015). These measurements have been used in further research to estimate other dinoflagellate species genome sizes (Cuadrado *et al.* 2019). Although this method of measuring genome size is standard, other characteristics of the dinoflagellate genome may lead to discrepancies that could potentially lead to overestimations of the genome size with this method, such as their 1) condensed chromosomes throughout the cell cycle, 2) lack of conventional

histones and nucleosomes, and 3) presence of modified bases. These potentially dubious estimations may be leading to overestimation of dinoflagellate genome sizes.

Unlike other eukaryotes, dinoflagellate nuclei contain chromosomes that remain condensed throughout the cell cycle in a liquid crystalline state (Cuadrado *et al.* 2019; Gornik *et al.* 2019; Wong 2019). Interestingly, the most recent phylogenies reveal the last common ancestor of the alveolates had a typical nucleus, leading to the idea that the dinokaryon has evolved into its unique biochemistry as a novel development unique to the dinoflagellate clade (Gornik *et al.* 2019). The crystalline structure of these chromosomes is known to cause strong birefringence, due to phase differences between light rays that pass through the densely packed DNA (Chow *et al.* 2010; Cuadrado *et al.* 2019). This birefringence may interfere with fluorescence measurements by altering the path and intensity of emitted light, potentially leading to inaccuracies in quantifying DNA content (Wong 2019).

The presence of crystallized and condensed chromosomes through the cell cycle as characteristic of the dinokaryon is theorized to be due to the lack of conventional histones and therefore the absence of nucleosomes (Cuadrado *et al.* 2019; Wong 2019). Histones are basic structural units of eukaryotic chromosomes and evolutionary extremely well-conserved as organization components of the eukaryotic nucleus (Kornberg 1974; Mariño-Ramírez *et al.* 2006; Oudet *et al.* 1975). Typically, eukaryotes display a nuclear DNA to protein ratio of 1:1, which typically means that the mass of histones to DNA is equal (Shechter & Allis 2007). Due to the lack of histones, dinoflagellates tend to skew towards much higher DNA:protein content, with a ratio of 10:1 (Hackett *et al.* 2005; Lin 2024). Conventional DNA staining may be impacted by this difference as the lack of nucleosomes may lead to non-linear staining, especially in the case of there being increased DNA dye binding sites compared to histone-compacted DNA. In lieu of histones, dinoflagellates have instead acquired other proteins, including dinoflagellate/viral nucleoprotein

(DVNP), major basic nuclear protein (MBNP) often also called histone-like protein (HLP) or bacterial HU-like protein (HCc) (Lin 2024). The MBNPs are similar in mass to histones, but do not form the octameric complexes like histones to create nucleosomes. Instead, MBNPs attract DNA with highly basic pH with lysine residues and divalent cations like Ca^{2+} and Mg^{2+}, therefore just condensing the DNA without creating complex structures (Kato *et al.* 1997). Syndineans typically possess DVNPs but lack MBNP. They also have notably small genome sizes, sometimes under 0.1 Gbp, exemplified by the parasitic *Amoebophrya* lineage (Bachvaroff 2019; Gómez 2012; Lin 2024). In contrast, "core dinoflagellates" feature both DVNP and MBNP—the latter consolidating several DNA-binding proteins including histone-like proteins (HLPs).

Another unique feature of the dinoflagellate genome is modified bases, particularly 5-hydroxymethyluracil (5-hmU) in place of thymine (Rae 1973). The 5-hmU replaces a significant amount of thymine in dinoflagellate genomes and was first discovered in 1973 in *Gyrodinium cohnii* (now known as *Crypthecodinium cohnii*) (Rae 1973). Since its discovery, it has been quantified from 12-68% of dinoflagellate genomic thymine bases, depending on the species (Lai *et al.* 2025; Marinov *et al.* 2024; McEwan *et al.* 2008). This modified base is ubiquitous in the dinoflagellate genome and is incorporated during DNA synthesis, rather than as a post-synthetic modification (Galleron 1984). Recent research has shown that 5hm-U may be more heavily enriched in repetitive regions of the genome, and may be introduced after synthesis at specific sites, which are also common in dinoflagellate genomes (Li *et al.* 2024b; Marinov *et al.* 2021). The substitution of this base may affect genome estimations, as it may influence dye accessibility and fluorescence intensity.

Dinoflagellate chromosome counts have ranged from as little as 7 (typically in symbiotic or parasitic species) to as many as 91, ranging from 16kb up to 120Mb each (Lin 2024). This range displays not only a very high chromosome count, but also some of the largest for eukaryotic

genomes. In comparison to *Arabidopsis thaliana*, which has only 5 pairs of chromosomes, dinoflagellate species have a much larger number (Lysak *et al.* 2006). The number of chromosomes in a genome increase the complexity, making assembly of discontinuous regions more error-prone (Hjelmen 2024). Scaffolding efforts also become more complex to account for the orientation and mapping of the reads along the chromosomes. As chromosome numbers increase, typically proportion of repetitive regions increase as well due the increase in heterochromatic regions.

Not only are dinoflagellate genomes estimated to be quite large, but many previous assembly attempts have run into the large amounts of repetitive content. Repeats cause ambiguity in read placements, often causing collapsing of repeat regions (Morton *et al.* 2020). High amounts of repetitive content can also lead to fragmented assemblies, or assemblies that are misassembled due to shared repetitive regions between non-repetitive segments. Repetitive content of genomes can be attributed to multiple types of elements. A commonly significant proportion of repetitive elements come in the form of transposable elements – these are mobile sections of DNA that can change their position in the genome, causing structural and functional changes (Fueyo *et al.* 2022). Satellite DNA makes up another category of repetitive elements, which are tandemly repeated, non-coding sequences, typically located in transcriptionally inactive heterochromatic regions of genomes (Thakur *et al.* 2021). Tandem gene arrays are also very common in dinoflagellates, with copy numbers being estimated into the thousands for some cases (Bachvaroff & Place 2008; Beauchemin *et al.* 2012; Stephens *et al.* 2020). These tandem gene repeat regions can also be hard to elucidate using short-read assembly.

Current dinoflagellate genome assemblies range from 55–70% repeat content (Allen *et al.* 1975; González-Pech *et al.* 2021; Jaeckisch *et al.* 2011; Li *et al.* 2024b; Stephens *et al.* 2020). Mammalian models such as humans and mice have 50-69% and 45% repetitive content,

respectively (Biémont 2010; de Koning *et al.* 2011), while Drosophila is closer to 35%. *Arabidopsis*, a common plant model, is known to have 14-27% repetitive content in its genome. The comparison to model genomes shows that dinoflagellate have relatively high ratios of repetitive content in their genomes compared to other researched organisms, making genomic assemblies a challenge.

Lastly, dinoflagellate genomes are difficult to annotate due to genes being relatively distant from model annotations, unique intron-exon boundaries, and dramatic differences in gene compactness, duplication, and assembly, making reference assemblies notably under-annotated, with many predicted genes lacking functional characterization (Lin 2024). Current dinoflagellate genome assemblies host over 50% predicted genes of unknown function. Of the predicted genes in dinoflagellate genomes, many have been categorized as 'dark' genes, and have no significant similarity to known proteins, although they can remain conserved within the dinoflagellate lineage (Stephens *et al.* 2018). This lack of annotation for millions of potential gene products displays a high potential for dinoflagellate genomes to possess many novel proteins with potentially unique functions. Advancements in functional genomics tools and more comprehensive databases are needed to improve the annotation and characterization of these enigmatic genes.

Gene Expression in Dinoflagellates

Introduction to Expression

Differential gene expression analysis displays shifts in gene expression under varying conditions utilizing RNA-sequencing data (Rosati *et al.* 2024). Differential gene expression analyses under various conditions in dinoflagellates have consistently reported low numbers of differentially expressed genes (DEGs) along with narrow fold-changes, but results have been conflicting with various methods and interpretations (Zaheri *et al.* 2019). For example, a 2019 study of

transcriptional responses to light in *Symbiodinium* found very minimal differential expression using multiple methods of measurement including northern blots, qPCR, and RNA-seq analysis (Zaheri *et al.* 2019). In this case the authors theorized that many analysis pipelines may present opportunities for false positives, especially with large datasets such as RNA-seq. Paradoxically, studies of bioluminescence clearly demonstrate differences in protein levels between day and night. Studies on *Pyrocystis*, *Karenia*, and other *Symbiodinium* species show 0.1 to 3% DEGs at Log-Fold Change (logFC) around 2 across conditions, although this is still quite low compared to genome-wide studies in model species that range from 5-10% global differential expression in different conditions (Baumgarten *et al.* 2013; Girardot *et al.* 2004; Lidie *et al.* 2005; Okamoto & Hastings 2003; Seok *et al.* 2023; Van Dolah *et al.* 2007; Xiang *et al.* 2015). Studies exposing *Karlodinium*, *Levanderina,* and *Prorocentrum* to algicides and heavy metals showed up to 20.6% DEGs, although these treatments coincided with mass cell-death where relative *in vivo* fluorescence of the cells dropped by 10% to 94%, insinuating potential degradation of cells and reduced RNA quality (Guo *et al.* 2016; Wang & Coyne 2024). In *Amphidinium* specifically, studies on differential expression are quite minimal, with only one study focusing on identifying DEGs in nutrient replete conditions, which claimed to observe differential expression between control and repleted conditions. This study involving nitrogen-starvation found that only 0.3% of transcripts showed significant levels of differential expression (Lauritano *et al.* 2017), although the culture conditions were not described after allowing the population to move to stationary phase in their assigned conditions before sampling. Lack of transcriptional response could be due to the unique chromosome structure in dinoflagellates, described above, with their lack of histones and permanently condensed chromatin (Zaheri *et al.* 2019). Due to this phenomenon, there is particular interest in how dinoflagellates may be regulating post-transcriptionally and which mechanisms may be increased in dinoflagellates, and which may be novel.

Spliced Leader (SL) Trans-Splicing

Dinoflagellate transcripts undergo spliced-leader (SL) trans-splicing, where a short exon from another RNA molecule, the SL RNA, is added to the 5' end of the pre-mRNA (Jones *et al.* 2024, no date; Roy & Morse 2013; Stover *et al.* 2006; Zhang *et al.* 2007b). This SL trans-splicing is a mRNA maturation process used across several distantly related eukaryotes. The SL sequence in dinoflagellates is 22 nucleotides, 5′-DCCGTAGCCATTTTGGCTCAAG-3′ (D = T, A, or G), and originates from a 65-66nt SL RNA (Jones *et al.* 2024; Roy & Morse 2013; Zhang *et al.* 2007b). This SL appears to be unique to the dinoflagellate nuclear transcripts and is not shared by organelle-encoded transcripts.

Polycistronic transcripts

The idea of dinoflagellates transcribing genes, such as tandem arrays, as polycistronic transcripts has mostly been debunked (Beauchemin *et al.* 2012). The closely related trypanosome group is known to use polycistronic transcription, transcribing long transcripts from a single promoter which contains transcripts to multiple products, which is then spliced (Jackson 2007). Polycistronic transcription in dinoflagellates was initially supported by the lack of orientation changes in dinoflagellate genes, but since then only evidence against this theory has been observed (Stephens *et al.* 2020). Most notably, in the Beauchemin *et al* 2012 study of *Lingulodinium* RNA-seq data, no polycistronic transcripts were observed. Not only this, but intergenic spacers between the tandemly arranged genes have not been well represented in the RNA pool.

Lack of Transcription Factors

A suite of transcription factors (TFs) is necessary in most known species systems in order to transcribe RNA, including 6 TF complexes and a TATA-binding protein (TBP) (Roy & Morse 2013). Dinoflagellate transcriptomes apparently lack identifiable versions of most of these TFs,

and so far, only TBP-like factors have been identified, although they contain an alternative TATA-interaction site that instead would potentially target <TTTT>. This is not unique within Alveolata as many genomes in this superphylum lack these common TFs.

RNA Editing

RNA editing is another phenomenon that affects dinoflagellate gene expression; this mechanism refers to changes in an RNA sequence after transcription where nucleotides can be switched out for another (Y. Chen *et al.* 2024; Zhang *et al.* 2024). RNA editing in dinoflagellates was first found in mitochondrial transcripts; here, nucleotide changes tend to be synonymous (or resolve stop codons) but impact the GC content of the transcript (Lin 2024). This was theorized to be due to the importing of tRNA from the cytoplasm into the mitochondria for translation, where the cytoplasmic tRNA may not function as well with the AT-rich mitochondrial genes. Since this discovery, RNA editing has also been observed in chloroplast genes and most recently in nuclear genes. In a 2023 study of *Durusdinium* and *Prorocentrum*, thousands of RNA editing sites were identified, with more being transitions than transversions, and a clear A-to-T bias (Dougan *et al.* 2023). Researchers have theorized that these changes to the RNA may be crucial for regulating gene expression in dinoflagellates.

mRNA Modifications

In most eukaryotic systems, N6-methyladenosine (m6A) is the most common internal mRNA modification, but in dinoflagellates N1-methyladenosine (m^1A) is observed as the most prevalent (Li *et al.* 2024a). In other eukaryotic systems, m^6A is recognized by m^6A-binding proteins and is responsible for initiation and progression of expression for many transcripts (Jiang *et al.* 2021). In dinoflagellates where m^1A dominates mRNA modifications, m^1A can block base-pairing, alter RNA structure, and affect protein-RNA interactions (Li *et al.* 2024a). Enrichment of m^1A appeared to be primarily in the 3'UTR and seems to be particularly involved in decreasing

translation efficiency. In *Amphidinium* studies, adenosine m^1A modification reached up to 3.05%, which was much higher than m^1A modification of around 1.65% in mammalian cells. Not only this, but m^6A modifications were only at 0.63% of adenosines in *Amphidinium*, which is similar to mammalian cells, and almost 5X higher than the m^6A content (Li *et al.* 2024a).

Summary

Understanding gene expression in dinoflagellates remains a pressing scientific challenge due to their uniquely complex and atypical genomic features. Unlike model eukaryotes, dinoflagellates exhibit persistently condensed chromatin, minimal transcriptional responses to environmental stimuli, and low proportions of differentially expressed genes even under stress conditions. Studies across species have reported DEGs ranging from less than 1% to around 20%, often complicated by cell death and RNA degradation, and analysis inconsistencies further cloud interpretation. In addition, dinoflagellates utilize distinct regulatory mechanisms, including SL trans-splicing, uncommon mRNA modifications such as m^1A, pervasive RNA editing, and a lack of conventional transcription factors. These post-transcriptional and epitranscriptomic strategies suggest that dinoflagellates may rely less on transcriptional control and more on RNA-level regulation to manage gene expression. Despite these intriguing features, key mechanisms remain poorly understood, particularly in underexplored genera like *Amphidinium*. Resolving these gaps is essential for decoding dinoflagellate biology and establishing a model system to better study the biology of this phylum.

What Defines a Model

"Model organisms" include species that are used to study biological processes that can be extrapolated to a wide range of organisms (Bertile *et al.* 2023; Leonelli & Ankeny 2013). Before an organism can be considered as a model, it likely first needs to be assessed as an "experimental

organism", which is a species chosen for experiments due to its suitability for a study into a particular biological process.

Dinoflagellate species are not viable 'models' due to their genomic complexity and lack of well-characterized molecular protocols to manipulate their expression patterns. But in order to understand the phylum Dinoflagellata more completely, specifically the photosynthetic clades, an experimental organism would need to be identified and possibly developed into a model organism for the phylum.

There are 4 parameters that model organisms possess (National Research Council 2000):

1. Ease of maintenance
2. Short generation times
3. Genetic tractability
4. Extensive research tools

These characteristics allow for ease of research as well as repeatability. Typically, model organisms are smaller and inexpensive to maintain. In the case of dinoflagellates, a model would need to be one that grows to a large density relatively quickly in culture, specifically in artificial seawater so that the growth characteristics can remain consistent. Short generation times also add to the ease of growth by allowing researchers to quickly observe multiple generations in a relatively short period of time.

What is Amphidinium?

Taxonomy

Amphidinium carterae is a species of dinoflagellate belonging to the family Amphidiniaceae, order Amphidiniales, and class Dinophyceae (Hulburt 1957; Murray *et al.* 2004). This species was originally described by Hulburt in Massachusetts in 1957 and has been placed in the "Operculatum" clade - one of two main lineage branches in *Amphidinium*—distinguished using LSU and ITS rRNA sequencing (Lorena María Durán-Riveroll *et al.* 2023). Cells are athecate (lacking cellulose thecal plates) and typically 10-17um long and 7-10um wide, dorsal-ventrally flattened, with a crescent-shaped highly-reduced epicone, large chloroplast, and central pyrenoid (Lorena María Durán-Riveroll *et al.* 2023; Gárate -Lizárraga 2012; Seoane *et al.* 2018).

Lifestyle

Amphidinium carterae is a cosmopolitan coastal species, typically residing in shallow marine and estuarine settings across various temperate and tropical zones. The species has been found globally from the North Atlantic to the Mediterranean, Golf of Mexico, Gulf of California, and New Zealand (Lorena María Durán-Riveroll *et al.* 2023; Lorena M. Durán-Riveroll *et al.* 2023; Gárate -Lizárraga 2012; Gárate-Lizárraga *et al.* 2019; Hulburt 1957; Pagliara & Caroppo 2012). *Amphidinium carterae* is primarily photoautotrophic and is a peridinin-pigmented dinoflagellate (Hofmann *et al.* 1996; Wagner *et al.* 2013). A*mphidinium carterae* primarily reproduces asexually through binary fission remaining motile through cytokinesis (Gárate -Lizárraga 2012; Murray *et al.* 2012). Rarely, *A. carterae* cells will reproduce sexually by producing gametes that fuse to form planozygotes – diploid cells that form cysts and have been shown to be triggered from environmental stressors.

Toxicity

Although *A. carterae* is not heavily associated with negative human health hazards, it has been associated with algal blooms and produces bioactive sterolysins called amphidinols (Cutignano *et al.* 2017; Lorena M. Durán-Riveroll *et al.* 2023; Hieda *et al.* 2021; Morales-Amador *et al.* 2021; Murray *et al.* 2012; Nuzzo *et al.* 2014). Amphidinols have a large 60-carbon open-chain structure that bind to cell membranes via sterol interactions causing permeabilization and hemolysis. Amphidinols have also been characterized as antifungal, antibacterial, cytotoxic, and ichthyotoxic (Hieda *et al.* 2021; Nuzzo *et al.* 2014).

Nuclear Features

Amphidinium carterae, like other core dinoflagellates, possesses a highly atypical nucleus known as a *dinokaryon*, characterized by perpetually condensed chromosomes, a relative paucity of canonical histones, and reliance on viral-derived or histone-like nuclear proteins to maintain chromatin structure (Galleron & Durrand 1979a; Lin 2024). In this species, genes are often arranged in tandem repeats, and its transcripts undergo trans-splicing, reflecting the unusual RNA processing pathways typical of dinokaryotic nuclei (Bachvaroff & Place 2008). Additionally, mitosis in *A. carterae* occurs via an extranuclear spindle system—microtubule bundles converge at multilobed vesiculate bodies adjacent to the nuclear envelope—underscoring the divergence from conventional eukaryotic mitotic mechanisms (Barlow & Triemer 1988). These nuclear characteristics align with broader attributes of dinoflagellate nuclei: enormous estimated genome sizes (ranging broadly from a few to hundreds of picograms), closed mitosis (dinomitosis) involving nuclear envelope invaginations, unusual chromosomal organization, and abundant non-histone nuclear proteins such as major basic nuclear proteins (MBNPs), histone-like proteins, and HU-like proteins that replace classical nucleosomal arrangements

Potential as a model dinoflagellate

Amphidinium carterae is at the very precipice of being regarded as a model for dinoflagellate research. This species has all the derived nuclear features, but not a truly massive estimated genome size, grows well in culture, produces a toxin, and has the standard pigmentation (Beedessee *et al.* 2020; Haq *et al.* 2017; Zhang *et al.* 2007a). As an unarmored, peridinin-containing dinoflagellate, *A. carterae* exhibits hallmark dinoflagellate traits and further research will uncover insight into plastid acquisition, genome evolution, and metabolic innovation in dinoflagellates, without the added obstacle of breaking through an armored exterior. Using *A. carterae* as a dinoflagellate model may also bridge the gap between heterotrophic deeply branching "protodinoflagellates" and core dinoflagellate (Kim & Kim 2007; Zhang *et al.* 2007a).

Amphidinium carterae CCMP1314 fits the parameters of a potential model organism for dinoflagellates. Firstly, it is one of the fastest growing photosynthetic dinoflagellate species, with a generation of time of as little as one day in sufficient conditions. It can also be easily maintained in culture for decades in enriched artificial seawater (ESAW) medium along with antibiotics (Berges *et al.* 2001; Liu *et al.* 2017).

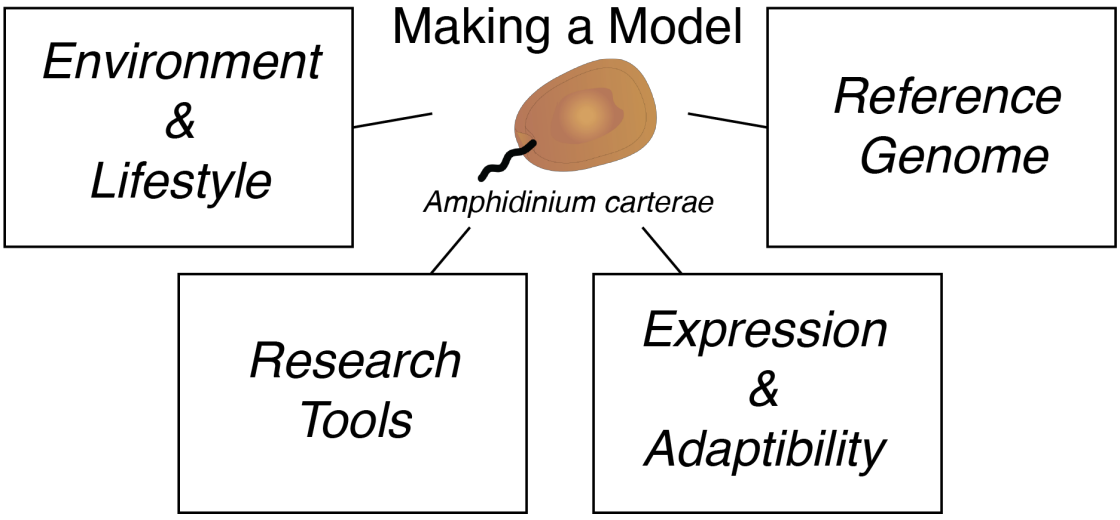
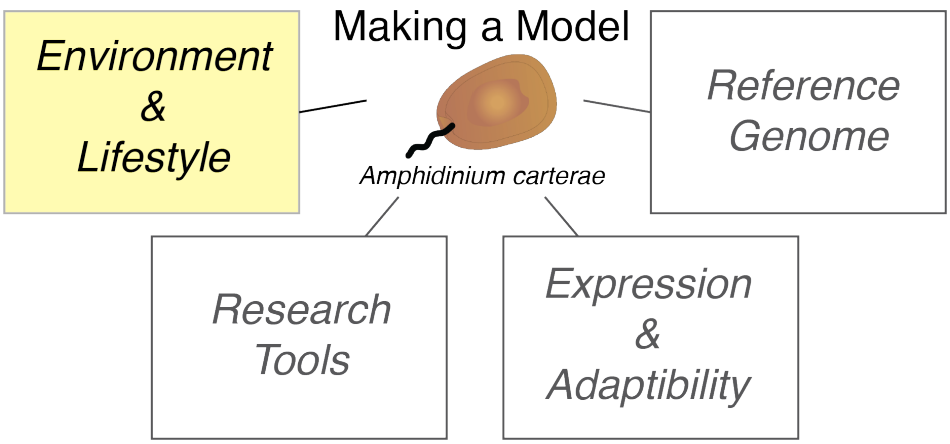


Figure i. Schematic of Research Areas in Establishing A. carterae as a Model.

To establish *A. carterae* as a model photosynthetic dinoflagellates species, research must be focused to elucidate its complex genome, and further genetic tools must be established to manipulate this organism for future scientific endeavors. The present dissertation outlines studies that aim to further these efforts by 1) exploring Environment & Lifestyle by evaluating culture conditions and characterizing the microbiome, 2) evaluating Research Tools such as genetic knockdowns, 3) analyzing Expression and Adaptability through diel/cell cycle measurements of physical traits and sequencing, and 3) creating a Reference Genome using novel long-read technology (Figure i).

Chapter 2: The Microbiome[1]



Environment and Making a Model

Characterizing an organism's microbiome is a critical step in establishing it as a model system because the microbial community plays a central role in host physiology, metabolism, immunity, and overall health. The composition and dynamics of the microbiome can influence experimental outcomes, from growth and development to stress responses, meaning that reproducibility and interpretation of results depend on understanding this underlying layer of biological variation. By defining the baseline microbial community, host-derived traits can be distinguished from microbiome-driven effects, symbiotic or pathogenic interactions can be identified, and environmental or laboratory studies have more realistic controls. Moreover, microbiome characterization provides a foundation for comparative studies across species and contexts, making the organism more broadly relevant as a model for studying host–microbe interactions, coevolution, and ecosystem-level processes.

[1] Results of this chapter are published in *Marine Drugs* under the title "Long-Read Sequencing Unlocks New Insights into the *Amphidinium carterae* Microbiome" (Judd et al. 2024).

Introduction

In the study of dinoflagellates, some species have shown to be culturable and are able to be kept in laboratory conditions in artificial seawater media. Many of these dinoflagellate cultures still contain abundant bacterial populations, likely originating from the original sample isolated from the environment (Alavi *et al.* 2001; Bolch *et al.* 2017; Green *et al.* 2004). The role of bacterial populations on dinoflagellate toxin production is one that remains hotly debated, with some prior studies showing evidence of dinoflagellate dependence on bacteria for toxin production and others showing the opposite (Albinsson *et al.* 2014; Bui *et al.* 2024; Moustafa *et al.* 2010; Stüken *et al.* 2011; Tarazona-Janampa *et al.* 2020; Williams *et al.* 2022). Some of these studies involve the filtering of bacteria from dinoflagellate cultures, followed by null amplification of 16S regions as verification to prove that a culture is axenic, and subsequent high-performance liquid chromatography (HPLC) analysis to show whether a toxin is still produced in the axenic culture (Hold *et al.* 2001; Lee *et al.* 2021; Su *et al.* 2007). The results of these studies have been contradictory.

Currently, culturing dinoflagellates with antibiotics is a common practice to select organisms of interest and maintain "axenic" conditions. Eukaryotic cultures can be maintained with antibiotics to decrease the effects of bacterial populations on downstream analysis such as sequencing efforts, toxin analysis, and translation rate studies (Liu *et al.* 2017). Although antibiotic use in dinoflagellate culturing can be useful, there is concern about bacterial antibiotic resistance allowing for regrowth of bacterial populations, as well as the impact antibiotics can have on the growth of the dinoflagellate culture.

We propose a new approach to investigate bacterial contributions to dinoflagellate cultures, using full assembly of all the bacteria present in the culture from long-read sequencing. These genomes can be used to identify pathways present in those bacterial populations, determine microbial

taxonomic composition, and estimate relative abundance (Sereika *et al.* 2022; Wick *et al.* 2023). Here, we use this approach in a toxin-producing *A. carterae* CCMP1314 culture with a long history of laboratory growth in the presence of carbenicillin, kanamycin A, and spectinomycin antibiotics. This study therefore focuses on the microbiome associated with this single strain that has been in culture for over 70 years as a case study for further microbiome studies involving long-read sequencing. This proposed approach offers a more comprehensive, unbiased measure of bacterial diversity and metabolism in *A. carterae* cultures, both with and without antibiotic treatment.

In this study, we found populations of bacteria in an *A. carterae* culture that thrived in the presence of antibiotics, identified antimicrobial genes within these populations, and observed how the dinoflagellate culture responds when reverted back to antibiotic-free growth conditions.

Methods

Cell Culturing

Amphidinium carterae CCMP 1314 was grown in ESAW (Enriched Water, Artificial Water) artificial marine media with a salinity of 20 supplemented with f/2 nutrients (75mg L^{-1} $NaNO_3$, 5mg L^{-1} NaH_2PO_4-H_2O, 1 mL trace metals [in 1 L: 10 mg $CuSO_4$, 22 mg $ZnSO_4$, 10mg $CoCl_2$, 180 mg $MnCl_2$-$4H_2O$, 6 mg Na_2MoO_4-$2H_2O$], and 1mL vitamins [in 1 L: 100 mg thiamine, 2 mg vitamin B12, 1mg biotin]) without silicates at a starting concentration of 10 K cells mL^{-1} and allowed to acclimate for 14 days in a 14:10 h light–dark period in 100 µmol photons m^{-2} s^{-1} at 25°C (Berges *et al.* 2001). Cultures of *A. carterae* were provided by the National Center for Marine algae and Microbiota, initially isolated in 1954 from Nantucket Sound. The cultures have been continuously grown with an antibiotic cocktail of 100 µg/mL carbenicillin, 50 µg/mL kanamycin A, and 50 µg mL^{-1} spectinomycin ("antibiotic-treated") for over a decade. For analysis, some cultures were weaned off growth in the presence of antibiotics for over a month

with consecutive re-growth in untreated media before subsequent DNA extraction ("antibiotic-free"). Cell counts were done using a Scepter 3.0 Cell Counter (Merck KGaA, Darmstadt, Germany) (Figure ii).

DNA Extraction and Long-Read Sequencing

For each dinoflagellate culture, the cells were spun down at 10,000 × *g* for 10 min to form cell pellets. The cell pellets were then resuspended in 0.1 M EDTA and 0.5% SDS with 200 µg proteinase K and allowed to incubate for 10 min at 50°C. Following this incubation, equal 16.6% volumes of 2% CTAB and 5 M NaCl were added, and once again the sample was allowed to incubate at 50°C for 10 min. After this, the solution was mixed 1:1 with chloroform and allowed to sit at room temperature for 10 min. The sample was then spun for 10 min at 10,000 × *g*, and the top aqueous layer was transferred to a new tube and mixed with a binding buffer from the DNA Clean & Concentrator kit (Zymo Research, Irvine, CA, USA). Aliquots of 750 µL were spun through the columns, followed by washing and finally elution in 20–40 µL of water. Small fragments were filtered out using a PacBio SRE/SRE-XL kit (Pacific Biosciences, Menlo Park, CA, USA) according to manufacturer's directions. A high-molecular-weight DNA library was then prepped using a Nanopore whole genome sequencing kit (SQK-LS114, Oxford Nanopore, Oxford, UK), followed by sequencing on a MinION, GridION, or PromethION device (Supp. Table i). The ONT data with the "super" accuracy Dorado basecalling model was used due to its recent increase in basecalling accuracy of 99.5% (Sereika *et al.* 2022).

Genome Assembly

Following sequencing, DNA data in pod5 files were basecalled by Oxford Nanopore's Dorado basecaller (7.2.13) using the dna_r10.4.1_e8.2_400bps_sup@v4.3.0 model. Sequences from the antibiotic-free culture runs or antibiotic-treated runs were pooled and assembled using the Flye *de*

novo assembler (2.9.1-b1780), filtering for sequences over 200 bases (Kolmogorov *et al.* 2019). Bacterial contigs were identified using BLASTn for 16S sequences from NCBI's 16S RefSeq or 16S microbial ribosomal databases. Bacterial matches from each assembly were identified, and duplicate matches between the antibiotic-free and antibiotic-treated runs were compared, and the most complete bacterial genomes were chosen for further analysis based on CheckM scores.

Phylogenetic Analysis

Once the bacterial genomes were identified, 16S sequences were extracted and identified using SILVA's Alignment, Classification, and Tree (ACT) service according to the global SILVA alignment for rRNA genes, as well as through NCBI's BLASTn feature (Medema *et al.* 2011). A consensus of the two was chosen based on match percentage. A phylogenetic tree was constructed using RaxML including 10 nearest neighbors identified by SILVA ACT from the database (Supp. Figure i).

Read Abundance Mapping

To quantify the relative abundance of bacterial genomes within the antibiotic-free and antibiotic-treated cultures, reads from each condition were aligned to the single, unified, reference set of bacterial contigs using minimap2, using ordinary minimizers as seeds (Li 2018). Samtools (version 1.16.1) was used to convert alignment SAM files to BAM, which were then sorted, indexed, and finally coverage and statistics files were created (Li *et al.* 2009). Genomes with greater than 90% coverage were considered present in the cultures, and abundance was calculated by the percent of reads mapping to that genome out of the total bacterial reads mapped in each condition.

Metabolic and Secondary Metabolites Analysis

Metabolic and biogeochemical functional trait profiles for each bacterial genome were assessed using METABOLIC-G using the KEGG database for pathway identification (Zhou *et al.* 2022).

Genome sequences were also input into antiSMASH for secondary metabolite biosynthetic gene clusters search (Medema *et al.* 2011).

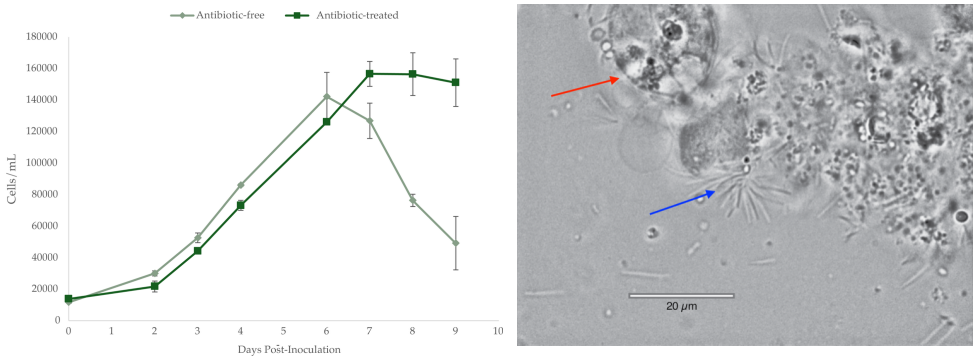
Antimicrobial Resistance Analysis

The Comprehensive Antibiotic Resistance Database (CARD) Resistance Gene Identifier (RGI version 6.0.3) software was used for the prediction of antibiotic resistance genes in the bacterial genomes (McArthur *et al.* 2013). Perfect, Strict, and Loose hits for genes were allowed, but partial genes were excluded. Protein predictions of antimicrobial resistance genes of interest were created using the AlphaFold Google DeepMind AI system (Jumper *et al.* 2021; Varadi *et al.* 2024).

Results

Bacterial Growth

The growth of *A. carterae* cultures both with and without antibiotics was found to be similar during both lag and log phases, with the daily maximum growth rate being about 63 ± 9% and 65 ± 9% for both treatments, respectively. Maximum cell densities were similar, with the antibiotic-free culture reaching 142,166 ± 15,382 cells/mL and the antibiotic-treated culture reaching 126,000 ± 1,311 cells/mL. Growth differed between antibiotic and control treatments once the stationary phase began. In the antibiotic-treated culture, cell growth stagnated, and density remained stable, while antibiotic-free cultures showed significant decreases in cell number and signs of lysing and death under microscopy (Figure ii-a). Cells within the antibiotic-free culture appeared to form aggregates, and dinoflagellate cells imaged at the beginning of the stationary phase in the antibiotic-free cultures displayed mass lysing and the appearance of rod-shaped bacteria overwhelming the culture, alluding to potential algicidal properties (Figure ii-b). These bacteria were not observed by microscopy in the antibiotic-treated cultures.



(a) **(b)**

Figure ii. *Amphidinium carterae growth with and without antibiotics.* ***(a)*** *Growth curves of A. carterae cultures;* ***(b)*** *image of lysed A. carterae (red arrow) taken at 100×, and the presence of rod-shaped bacteria (blue arrow).*

Bacterial Identification

Metagenome assembly of *A. carterae* cultures identified 15 bacterial contigs based on 16S rDNA identity, with genomes ranging from 2.9 to 6.0 Mb (Table i). There were 8 genomes assembled from the bacterized reads, 6 assembled from the antibiotic-treated reads, and 1 assembled from the combination of both read-sets. All of the genomes were assembled to some degree in all of the read set assemblies, with the prevalence of some species over others allowing for better assembly between the conditions. All 15 genomes were single contig genomes, and the Flye assembly reported 14 of the 15 being circular (all but *Oceanicaulis alexandrii*). The genome average depth of coverage ranged from 13 to 1612. Nearest neighbors of the 16S sequences found in these genomes were 98 to 100% identical to reference sequences in the Silva and GenBank databases (Supp. Figure i). The bacteria present included 10 Alphaproteobacteria, 1 Gammaproteobacteria, 3 Bacteroidia, and 1 Planctomycetota. Based on the very high (>98%) full-length 16S rDNA

identity to previously described species, the names of these high-identity 16S rDNA matches are used as an operational taxonomy for the remainder of the text.

Within the 15 genomes, between 2,890 and 5,758 genes were predicted; of these, an average of 62% ± 13% had functional assignments according to BV-BRC, and of the hypothetical proteins, 74% ± 21% were longer than 100 amino acids (Supp. Table ii). Genome quality was qualified as good for all of the genomes, with CheckM Completeness ranging from 77.5% to 100%, and 14 of the genomes showing above 90% completeness (all but the *Nitratireductor* sp.) (Parks *et al.* 2015). CheckM Contamination scores were below 5% for all the genomes as well.

Table i. Bacterial genome characteristics.

NCBI Sample Title [1]	**Length**	**AT%**
Labrenzia sp. strain ac12	6,053,707	40.9
Hoeflea alexandrii strain ac4	4,829,646	38.4
Algihabitans albus strain ac2	4,739,774	34.3
Ahrensia marina strain ac1	4,424,055	42.8
Seohaeicola saemankumensis strain ac14	4,275,009	36.6
Roseovarius mucosus strain ac13	3,969,289	38.9
Rhodophyticola porphyridii strain ac11	3,902,216	35.8
Oceaniradius stylonematis strain ac10	3,664,187	35.3
Oceanicaulis alexandrii strain ac9	2,992,841	37.7
Nitratireductor sp. strain ac15	2,917,504	39.8
Marinobacter adhaerens strain ac5	4,424,055	42.9
Cyclobacterium xiamenense strain ac3	5,806,256	51.6
Marivirga tractuosa strain ac6	4,787,102	65.3
Muricauda sp. strain ac8	4,366,883	58.0
Phycisphaeraceae SM1A02 strain ac7	3,415,114	34.9

[1] 16S sequence identification found using both the SILVA ACT service and the NCBI BLASTn feature (Quast *et al.* 2012).

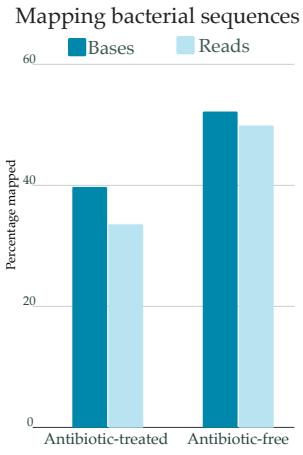
Bacterial Abundance

In the antibiotic-free cultures, 52.2% of the total bases and 49.9% of reads mapped to the 15 bacterial genomes, while in antibiotic-treated cultures, 39.8% of total bases and 33.5% of reads

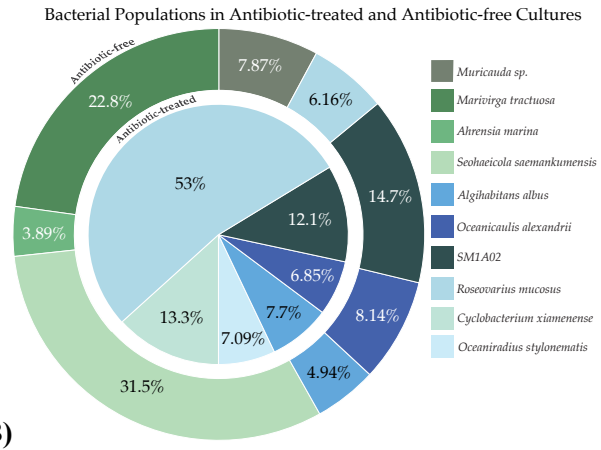
mapped to bacteria (Figure iii-a). Ten of the 15 bacterial genomes identified accounted for 90% of the total bacterial reads mapped in both conditions (Figure iii-b), and the remaining five taxa accounted for <10% of total bacterial reads.

All 15 bacteria were detectable at some level in all the batches of sequencing data, but the relative abundance differed substantially for 7 bacterial genomes. Under antibiotic treatment, *R. mucosus* was the most abundant single bacterium (at 53%) and then decreased to 6.16% without antibiotics. Without antibiotics, *Muricauda* sp., *Marivirga tractuosa, Ahrensia marina,* and *Seohaeicola saemankumensis* rose to the top 90% of the bacterial reads. In the antibiotic-treated culture, *O. stylonematis* and *C. xiamenense* populations were among the top 90% of bacterial reads, but these bacteria were not abundant without the antibiotic treatment.

The antibiotic-free and antibiotic-treated cultures both included 3 taxa at roughly similar proportions: SM1A02, *O. alexandrii*, and *A. albus*. When cultures were grown without antibiotics, the abundance of *SM1A02* and *O. alexandrii* slightly increased by 2.6% and 1.3%, respectively, while that of *A. albus* slightly decreased by 2.8%. The remaining 5 bacterial genomes were present at low levels in both culture conditions.



(A)



(B)

Figure iii. *Bacterial populations with and without antibiotics. (A) Percent of reads and bases mapping back to bacterial genomes from both the antibiotic-free and antibiotic-treated cultures, aligned using minimap2. (B) Bacterial abundance based on read mapping in the antibiotic-free and antibiotic-treated cultures.*

Bacterial Metabolic Pathways

The metabolic abilities of the bacterial genomes were assessed using the KEGG database, showing a wide variety of potential pathways (Figure iv, Supp. Table iii). Many of the bacterial genomes have predicted pathways for the production of amino acids, such as arginine, proline, serine, and threonine. Basic metabolic pathways such as carbohydrate metabolism and ATP synthesis were identified. Some of the bacterial genomes, including *R. mucosus, R. porphyridii, H. alexandrii, O. stylonematis, A. marina*, and *A. albus*, were found to contain predicted genes for photosystem II photosynthesis machinery, as well as some carbon-fixation pathways. The *Labrenzia* sp. genome, although the largest bacterial genome assembled, was also in low abundance in the antibiotic-treated culture and contained a part of the reductive phosphate pentose cycle and the Calvin cycle, but not photosystem II.

Nitrate-reduction pathways were found in 3 of the bacterial genomes, and denitrification machinery was identified only in the *M. tractuosa* genome. The *Nitratireductor sp.*, found in low abundance the antibiotic-treated culture, was negative for nitrate-reduction pathways despite the generic name, which aligns with prior research (Marasco *et al.* 2023).

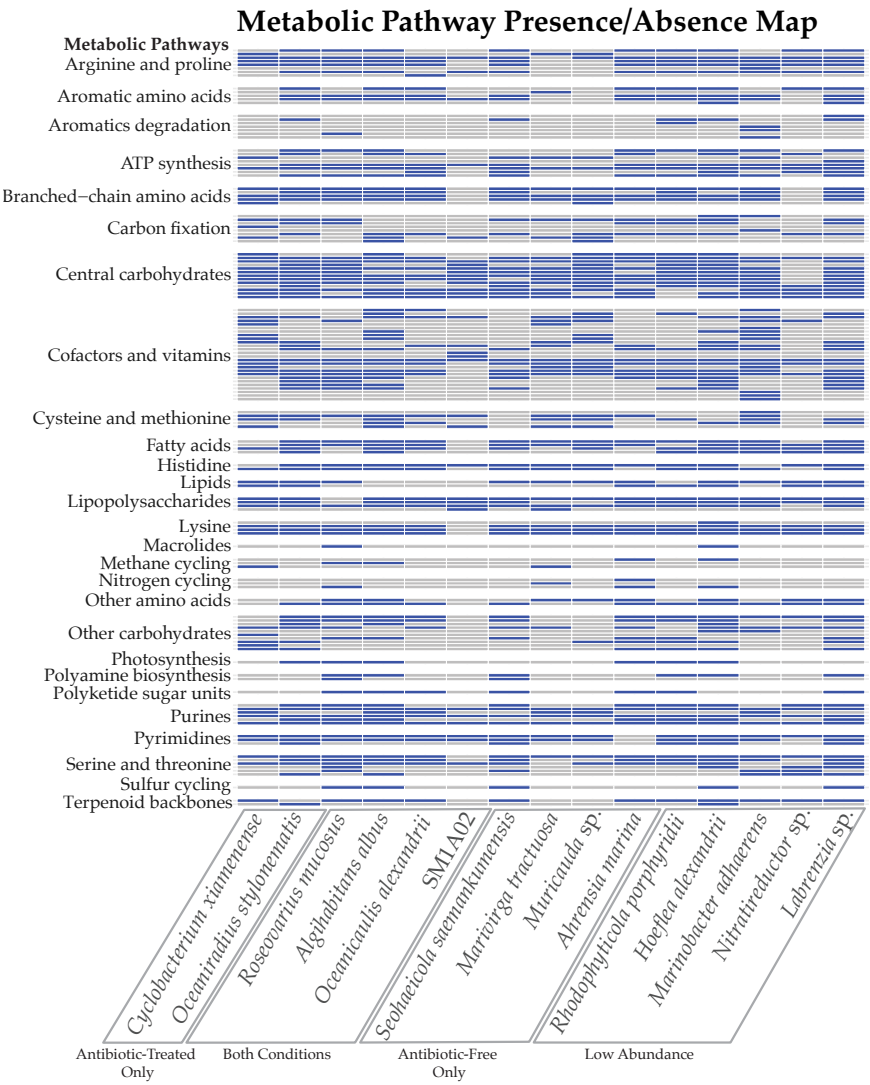


Figure iv. *Functional traits identified via CARD based on the KEGG database* (Zhou *et al.* 2022)*. Specific pathway names for each row are listed in Supplemental Table iii.*

Secondary Metabolite Synthesis

The assembled bacterial genomes were found to contain regions associated with the production of a variety of secondary metabolite compound classes, including building-block lactone structures such as butyrolactones, betalactones, and hserlactones (Figure v). Terpenes and aromatic compound synthase genes were the major natural product occurring amongst the genomes. Some polyketide synthases (PKSs) and non-ribosomal peptide synthases (NRPSs) were identified, but

none were predicted to participate in the biosynthesis of larger chemical structures according to antiSMASH, although they could potentially be part of a multi-species production.

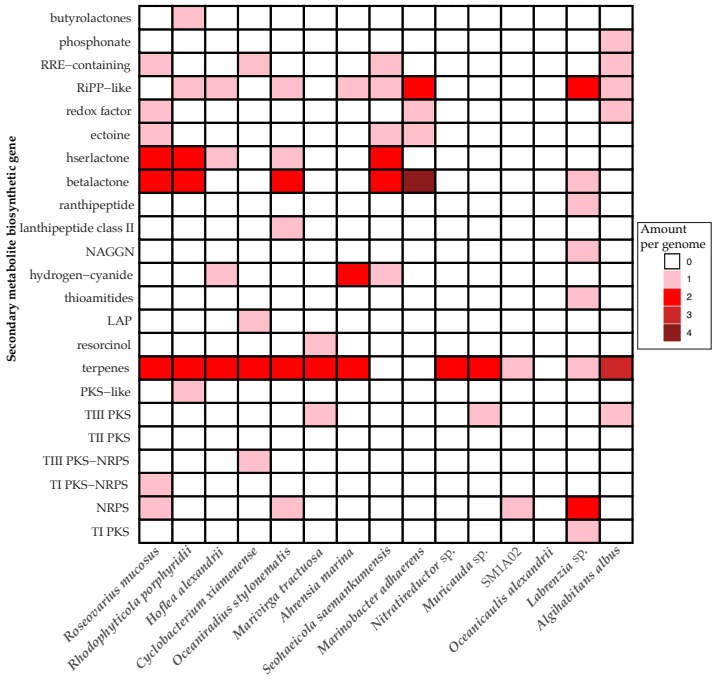
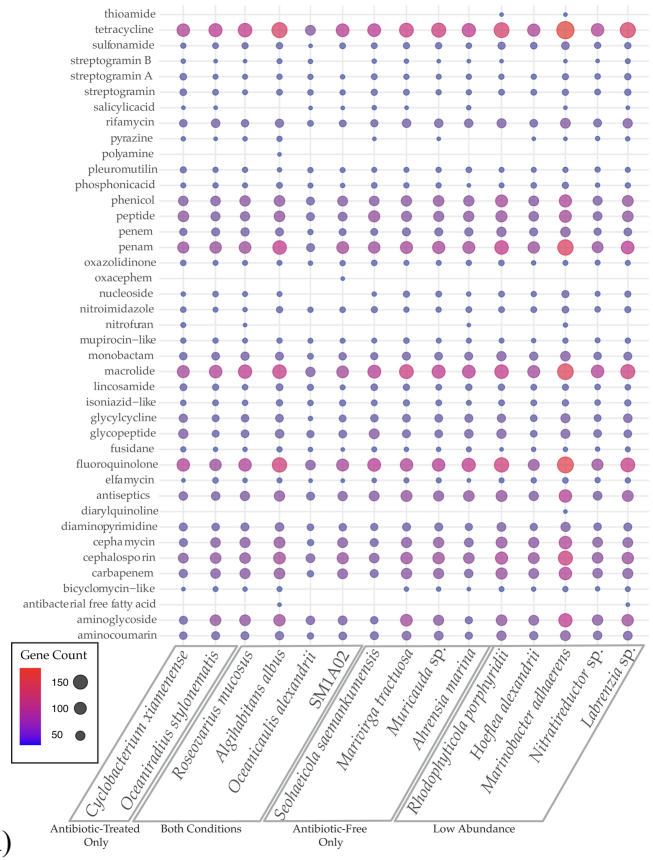


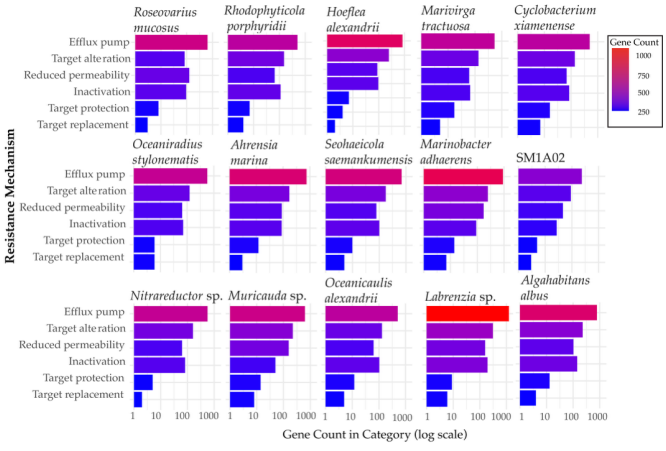
Figure v. *Secondary metabolites present in the bacterial genomes, identified by antiSMASH* (Medema *et al.* 2011).

Antibiotic Resistance

All of the bacterial genomes had antibiotic resistance genes in the CARD database search (Figure vi-A). Interestingly, all of the genomes had a large number of resistance genes to tetracycline, including *tet* genes for antibiotic efflux, ribosome protection, and antibiotic inactivation. The antibiotic resistance genes in the bacterial genomes found in the antibiotic-treated cultures did not appear to indicate a strong pattern of drug class resistance to the antibiotic cocktail used. Most of the antibiotic resistance genes were part of antibiotic efflux systems, especially in the assembled *Labrenzia* sp. genome (Figure vi-B). This was generally followed by mechanisms including drug-target alterations, reduction of drug permeability, drug inactivation, drug-target protection, and drug-target replacement, as identified in the CARD database.



(A)



(B)

Figure vi. *Antimicrobial resistance in assembled microbiome. (A) Antimicrobial resistance genes within each assembled bacterial genome (B) Specialty genes identified from the CARD database* (McArthur *et al.* 2013).

The dominant bacterial genome in the antibiotic-treated culture, *R. mucosus*, contained 13 resistance genes that were either completely unique to the taxa or found only within bacterial genomes found in the antibiotic-treated culture (Supp. Table iv). Of those 13, nine had the potential to confer resistance to carbenicillin, kanamycin A, and/or spectinomycin (Figure vii). Three of these genes encoded potential beta-lactamases, which inactivate beta-lactam antibiotics by hydrolyzing the peptide bond of the beta-lactam ring and may confer resistance to carbenicillin. One of the genes is similar to the efflux subunit of the AmrAB-OprM multidrug efflux complex, which impairs aminoglycoside accumulation. Lastly, five genes in the *R. mucosus* genome were identified as potential aminoglycoside acyltransferases, which may catalyze the AcCoA-dependent N-acetylation of amino groups on the aminoglycoside molecule (such as kanamycin A and spectinomycin).

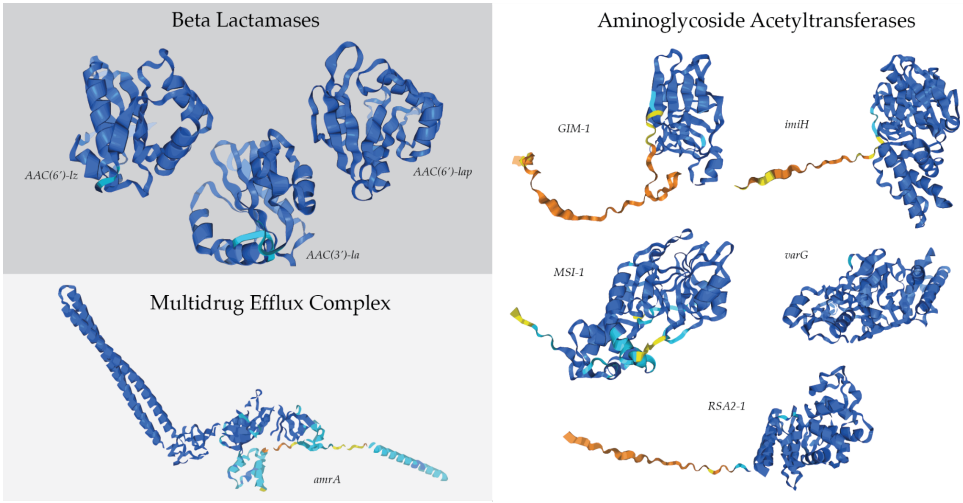


Figure vii. *Antibiotic resistance genes' protein structure predictions created by AlphaFold, unique to the assembled Roseovarius mucosus genome* (Jumper *et al.* 2021; Varadi *et al.* 2024). *Confidence in structure ranged based on per-residue model confidence score (pLDDT) from dark blue (pLDDT > 90), light blue (90 > pLDDT > 70), yellow (70 > pLDDT > 50), to orange (pLDDT < 50).*

Discussion

Among the 15 assembled genomes, every genome had a high (>95%) identity 16S rDNA match to an existing annotated sequence, suggesting that the culture did not contain any entirely novel bacteria. Many of the assembled genomes were identified as bacterial species that have previously been observed as co-existing with dinoflagellates. Interestingly, the assembled Planctomycetota bacteria were identified as SM1A02, an uncultured strain associated with many dinoflagellate cultures (Rambo *et al.* 2020). This species' genome has previously been reconstructed using metagenomic assembly and binning. Research by Baker *et al.* found two of these SM1A0*2* genomes to be 2.6 and 2.9 Mb, while we have assembled a 3.4 Mb genome here. Most recently, an assembly of SM1A02 was produced from a *Karlodinium* culture, which agrees with our genome length of 3.4 Mb (Tizabi *et al.* 2025). SM1A02 is thought to likely be an anammox bacteria—efficient at nutrient removal, specifically through anaerobic ammonium oxidation (Vico *et al.* 2021). This ability to oxidize ammonium to nitrogen gas may have an impact on the close association with dinoflagellate species.

Roseovarius mucosus was also identified, similar to a species found with the dinoflagellate *Alexandrium ostenfeldii*. In prior research, genes in *R. mucosus* were found that may play crucial roles in the interrelationship of the bacterium and dinoflagellate, such as genes for dimethyl sulfoniopropionate (DMSP) utilization. Research on the close interactions of DMSP-degrading *Roseobacter* species with DMSP-producing dinoflagellates is well-established (Miller & Belas 2004). Our metabolic results show that *R. mucosus* likely has pathways for thiosulfate oxidation, supporting these past findings (Ding *et al.* 2023). The ability for thiosulfate oxidation may also have a connection to the common sulfation of toxic amphidinol products produced by *A. carterae* or more broadly for sulfur cycling within the cultures (Haq *et al.* 2023; Nuzzo *et al.* 2014). The pathway for assimilatory nitrate reduction found in the *R. mucosus* genome has been observed as a potential nitrogen source by some (LeKieffre *et al.* 2020; Olofsson *et al.* 2019). The pathway

for polyamine biosynthesis for putrescine and spermidine has also been found in the *R. mucosus* genome, which may play an essential role in dinoflagellate growth (Liu *et al.* 2019; Nishibori & Imai 2013; Wang & Coyne 2024). The presence of PSII and other carbon-fixing pathways also aligns well with prior research of this species (Biebl *et al.* 2005).

Hoeflea alexandrii and *Oceanicaulis alexandrii* were discovered with a dinoflagellate species of *Alexandrium* as well as bacterial species *Labrenzia alexandrii* and *Nitratireductor alexandrii*, which have high 16S identity to the *Labrenzia* and *Nitratireductor* species assembled here (Biebl *et al.* 2007; Jiang *et al.* 2020; Palacios *et al.* 2006; Strompl 2003). *Marinobacter adhaerens* has been found in close association with *Pyrodinium*, another toxin-producing dinoflagellate (Chin *et al.* 2023). *Muricauda* species have been previously associated with *Amphidinium* as well (Chen *et al.* 2019). A *Seohaeicola* species' genome was recently assembled from a culture of *Karlodinium*, another toxin-producing dinoflagellate species (Tizabi *et al.* 2025). Some species, such as *O. stylonematis*, *A. albus, C. xiamenense* and *R. porphyridii*, were discovered in association with other kinds of microalgae, such as diatoms and red algae (Chen *et al.* 2014; Chernikova *et al.* 2020; Jeong *et al.* 2019; Jung *et al.* 2019; Wang *et al.* 2019).

Some species found only in the antibiotic-free cultures may be opportunistic due to the nutrient availability during dinoflagellate senescence. Both the assembled *R. porphyridii* genome and the *H. alexandrii* genome were found in very low abundances in the antibiotic-free culture. The *R. porphyridii* species is a genus of purple non-sulfur bacteria known to be halophilic and have the ability to perform photosynthesis (Jung *et al.* 2019). Thiosulfate oxidation pathway genes were also found in the *H. alexandrii* genome.

The transition from antibiotic treatment to untreated is unlikely to have resulted in the introduction of new bacteria to the culture, as all taxa were seen and could be at least partially assembled in sequencing of either antibiotic-treated or untreated cultures. Thus, the microbiome

shift is more likely due to stronger growth of some species over others when antibiotics were present or absent rather than the recruitment of novel species during culture changes, which were performed under sterile conditions.

Previous studies have demonstrated that the use of the KAS-antibiotic treatment (kanamycin A, ampicillin, and streptomycin) can be used to favor pigmented bacterial species (Takagi *et al.* 2023). Similar mechanisms may be why we saw such a shift towards bacterial populations with PSII and carbon fixation systems with the use of our antibiotic cocktail, from 12.5% to 50% of the highly abundant bacterial species having PSII pathways with the antibiotic-treatment. The *R. mucosus* identified in past dinoflagellate cultures, which dominated the antibiotic-treated microbiome in this experiment, has been shown to contain bacteriochlorophyll *a* (Biebl *et al.* 2005). There is also evidence of dinoflagellates protecting certain pigmented bacterial populations from antibiotics, as the pigmented bacteria may be protecting the microalgal cells from light stress via carotenoid production, which has previously been shown to be produced by multiple assembled bacterial species (Figure v) (Takagi *et al.* 2023). In prior research regarding the coral symbiont dinoflagellate *Symbiodinium*, the bacterial microbiome was observed to support the dinoflagellate's PSII yield and decrease the production of reactive oxygen species (ROSs) (Motone *et al.* 2020).

Antimicrobial resistance appears to be generally common amongst bacterial populations found in dinoflagellate cultures. In the case of the assembled genomes here, all had hundreds of potential antibiotic resistance genes that likely allowed their broad prevalence (Figure vi). The reasons why some bacteria were found to survive better in antibiotic-treated or antibiotic-free conditions could be due to multiple causes. One may be that the minimum inhibitory concentration of the antibiotic used may differ from the actual concentration tested (Kowalska-Krochmal & Dudek-Wicher 2021). This may be due to the mechanism of resistance, or the antibiotic's resistance to

degradation (such as in the case of carbenicillin compared to ampicillin) (Rolinson 1998). The specific genes found within each of these resistance groups may have varying efficacy against the antibiotics as well. In the case of *R. mucosus*, which best endured the antibiotic treatment used here, it may be that one of the nine antibiotic resistance genes had greater efficacy over one or more of the antibiotics used compared to the machinery found in the other genomes (Figure vii).

The extent to which the bacteria from the assembled genomes are mutual, commensal, or deleterious to the *A. carterae* population is still obscure. The fact that the dinoflagellate population significantly decreased and showed signs of mass lysis without antibiotics is circumstantial evidence that at least one abundant bacterial species in the antibiotic-free culture could be the cause. Prior research on antibiotic effects on dinoflagellate growth has shown various results. In some cases, dinoflagellates appear to require their associated microbiomes to survive (Bolch *et al.* 2017, 2011). In the case of the antibiotic cocktail used here, the growth results align with previous observations of the antibiotic-treated *Amphidinium* cultures having a slightly extended growth phase and the ability to maintain higher densities of dinoflagellate cells (Liu *et al.* 2017). Based on microscopy analysis, we found 10 µm length rod-shaped bacteria that began to accumulate around the start of mass cell lysis which could also potentially be suspects that harm the dinoflagellates and may possess some algicidal properties, although this cannot currently be confirmed. This culprit may be *Marivirga tractuosa* or *Seohaeicola saemankumensis* due to the increased abundance found in the antibiotic-free culture (Figure iii-B), as well as prior descriptions of this species as being rod-shaped. Cells of *M. tractuosa* can be between 10 and 50 µm long, while *S. saemankumensis* has been shown to be up to 5 µm in length (Nedashkovskaya *et al.* 2010; Pagani *et al.* 2011; Yoon *et al.* 2009). The *M. tractuosa* genome lacks many main metabolic pathways, such as amino acid biosynthesis, suggesting that this species requires resources gained from the lysed dinoflagellate cells, although they may not directly harm the dinoflagellate cells. The *M. tractuosa* genome has also been found to have complete

denitrification pathways, which may contribute to a loss of bioavailable nitrogen in the culture (Dagenais-Bellefeuille & Morse 2013). The Coenzyme M pathway alludes to methanogenic abilities and production of methane, and potentially the use of dinoflagellate-released DMSP as a precursor (Damm *et al.* 2008; Lin *et al.* 2021). Harmful algal blooms have been observed to precede methane increases in aquatic environments, which may be in part due to the shift in the microbial community.

Prior research has been contentious over the secondary metabolite synthesis potential of dinoflagellate microbiomes and contrasting results have identified toxin production as a product either of the bacterial community or the dinoflagellate cells themselves (Albinsson *et al.* 2014; Bui *et al.* 2024; Monroe *et al.* 2010; Moustafa *et al.* 2010; Stüken *et al.* 2011; Tarazona-Janampa *et al.* 2020; Tse *et al.* 2020; Uribe & Espejo 2003; Wang *et al.* 2018). The 15 apparently complete genomes assembled from this culture likely represent the bulk of prokaryotic diversity due to the extent of our sequencing and the production of full, well-covered, complete genomes across a wide range of sequencing coverage and depth. Any missing diversity would likely be in very low abundance to evade detection and is not likely to be present at a level to contribute to toxin biosynthesis. Similarly, gene annotation provides a potentially complete picture of the culture metabolic potential. However, a large fraction of predicted genes was unannotated, likely due to imperfect prediction of protein coding genes as well as knowledge gaps of every possible bacterial pathway. Our genome analysis into secondary metabolite synthesis has shown no evidence of potential full bacterial origin for a processive multidomain PKS gene responsible for the amphidinol toxins associated with our *A. carterae* culture, although we cannot rule out the potential of the bacterial producing toxin precursors. Since multidomain, processive PKS genes are very large open reading frames rich in easily defined conserved domains, these genes would be unlikely to have been missed in the genomes described here. Several multidomain PKS/NRPS genes derived from bacteria are present broadly across core dinoflagellate transcriptomes, which

generally express a surfeit of domains associated with toxin production and lipid synthesis whether or not they are documented toxin-producing species (Williams *et al.* 2022). More research could be done to see what effect metabolic pathways, such as thiosulfate oxidation, may have on toxin production. Although it seems most likely that *A. carterae* independently synthesizes amphidinols, the bacterial populations may contribute resources for the task, such as acetate (Haq *et al.* 2023; Houdai *et al.* 2001; Rein & Borrone 1999). Of the bacterial genomes assembled, only *M. adhaerens* was shown to have a complete pathway for a phosphate acetyltransferase−acetate kinase pathway, which produces acetate from acetyl-CoA (Figure iv). *Oceaniradius stylonematis, S. saemankumensis, R. porphyridii, H. alexandrii,* and the *Labrenzia* sp. all had complete phenylacetate degradation pathways to produce acetyl−CoA, which may serve as a precursor to acetate synthesis, and *O. stylonematis, S. saemankumensis, R. porphyridii, R. mucosus, A. albus, O. alexandrii, M. adhaerens,* and the *Labrenzia* sp. had complete pathways for leucine degradation to acetyl−CoA as well.

The microbiome of algal species has been shown to contribute necessary vitamins and products to dinoflagellate species, the most recognized being cyanocobalamin (B12) (Cruz-López & Maske 2016). Based on our metabolic findings, the introduction of B vitamins into dinoflagellate growth media does not appear to be redundant with the biosynthetic abilities of the microbiome. The vitamins added to our ESAW media preparation include biotin (H), B12, and thiamine (B1) to increase growth rate and final yield (Berges *et al.* 2001). Of the bacterial genomes assembled, only the *M. adhaerens* genome had a pathway identified to synthesize biotin, and this species was in very low abundance. The only highly abundant species with aerobic and anaerobic pathways for the synthesis of B12 were *R. mucosus* and *O. stylonematis*, which were significantly more abundant in the antibiotic-treated cultures, and their decline without antibiotics may have been a factor in the cell mortality of *Amphidinium* as the nutrients in the culture diminished over the log phase. Only *A. albus* within the highly abundant bacterial species had a pathway for vitamin B1

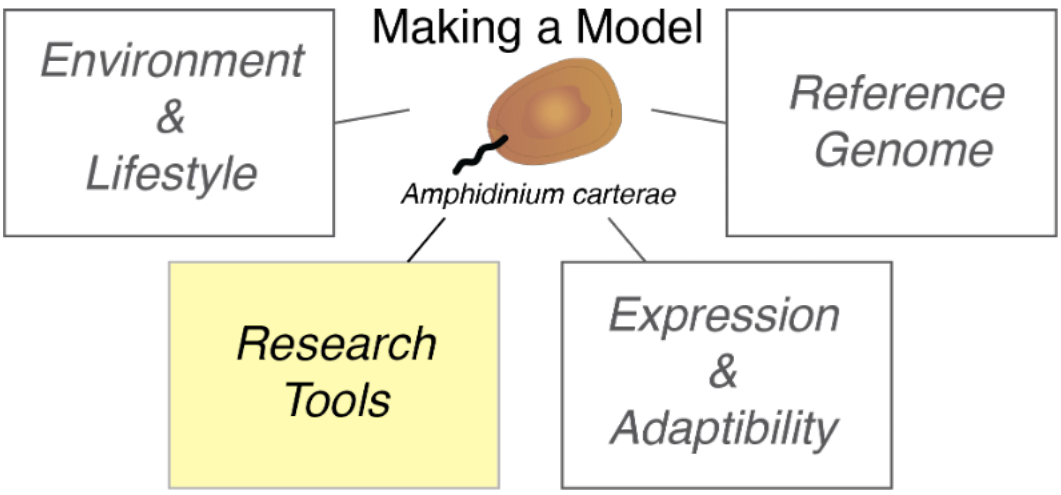
synthesis, and yet this was only through a salvage pathway that utilizes precursors or similar compounds in the surrounding media for B1 biosynthesis (Gonçalves & Gonçalves 2019).

The Planctomycetota bacteria SM1A02 is the only assembled genome in the *Amphidinium* culture to have full pathways for menaquinone (vitamin K2) biosynthesis. Although vitamin K1 is conventionally known as a redox cofactor in plants and green algae, vitamin K2 can also be a secondary electron acceptor of PSI in some algal and archaeal species (Del Mondo *et al.* 2020). Vitamin K2 can also shuttle electrons between different respiratory complexes in anaerobic respiration or aerobic respiration in a microaerophilic environment (Cenci *et al.* 2018). The effect of vitamin K2 bioavailability for dinoflagellate species remains to be seen.

Conclusions

Here long-read sequencing was used to qualify the microbiome surrounding a dinoflagellate culture. With the extensive use of long-read sequencing, our results showed that we were successful in assembling full bacterial genomes that could be used to assess the functional qualities of the associated bacterial species. Although *A. carterae* may be grown without antibiotics, the viability of the culture and subsequent sequencing analyses may benefit from their use. The observed die-off in the *A. carterae* cultures during the end of the log phase without antibiotics present may have been due to the proliferation of one of the more antibiotic-sensitive bacteria identified. We also found a decrease in the proportion of bacterial reads sequenced with the use of antibiotics. As novel sequencing technologies such as efficient long-read sequencing become more readily available, techniques to target sequences of interest will allow for more productive sequencing efforts. The data suggest that both comprehensive and quantitative microbial genome sequencing can be accomplished from this culture, which could in the future be expanded to work with in situ sequencing of dinoflagellate blooms.

Chapter 3: A Strategy for Genetic Knockdown[2]



Introduction

The establishment of a model organism requires the verification of various genetic research tools in order to study that species. Studies on dinoflagellate gene expression has been hampered by their unusual cell biology (Jones *et al.* 2015). Their genomes are estimated to be larger than typical protists, with an average of 1.2–112 × 10^9 base pairs of DNA per haploid genome (Verma *et al.* 2019), whereas other protist genomes are often approximately 10^8 base pairs in size (Brayton *et al.* 2007; Ebenezer *et al.* 2019). Dinoflagellate chromosomes are condensed into liquid crystalline states throughout the cell cycle and lack nucleosomes, instead using histone-like proteins (HLPs) that are more similar to bacterial DNA binding proteins. Many dinoflagellate genes are organized in multiple copies as tandem repeats. Increasingly transcriptomic data has shown that dinoflagellates express numerous genes, yet over 50% have no match to known sequences (Zhang & Lin 2019). The function of these sequences, as well as the effects of identified sequences, still need to be established through functional genomic studies.

[2] Results of this chapter are published in *Microorganisms* under the title "A Strategy for Gene Knockdown in Dinoflagellates" (Judd & Place 2022).

Here, an antisense-based knockdown approach was used in order to study how a decrease in target gene expression effects dinoflagellate metabolism. In prior research, the introduction of antisense-oligomers to dinoflagellate cells has been hampered by their thick, cellulosic cell wall (Kwok *et al.* 2007; Yan *et al.* 2020). Other studies have bypassed this obstacle by first preparing spheroplasts, cells with a completely or partially removed cell-wall (Chan *et al.* 2019; Kwok *et al.* 2007; Yan *et al.* 2020). Spheroplast production is done by incubating cells on plates in a polyethylene glycol (PEG) solution, which promotes fusion of the vesicles and cell membrane, and ultimately a decrease in total cellulose. So far only two studies have been successful in achieving gene knockdown with this spheroplast procedure; targeting a condensin subunit and targeting a cellulose synthase (Chan *et al.* 2019; Kwok *et al.* 2007; Yan *et al.* 2020). Once introduced, the antisense-oligo was able to bind to cytoplasmic mRNA and knockdown expression of the target gene. Although gene expression could be quantified in this way, it appears that some physiological effects were hidden by the effects of PEG on the cell wall, which causes the cells to lose rigidity. Also, the need for cell plating, rather than cell culturing, immensely limits the species of dinoflagellates that can be studied since many will not grow outside of a liquid medium.

There has also been evidence of RNA interference (RNAi) machinery within dinoflagellates, a naturally occurring mechanism for gene silencing through various methods such as RNA degradation, transcriptional repression and translation inhibition (Akbar *et al.* 2018a; Baumgarten *et al.* 2013; Zhang & Lin 2019). One study observed the effects of RNAi silencing tool on the proton-pump rhodopsin and CO_2-fixing enzyme Rubisco encoding genes in dinoflagellates by introducing small interfering RNAs (siRNAs) to dinoflagellate cultures via immersion. Results showed success in gene suppression within the two dinoflagellate species studied, *Prorocentrum donghaiense* and *Karlodinium veneficum* (Zhang & Lin 2019). This decrease in gene expression was observed with a decrease on overall growth rate for both species as well, compared to the

control green fluorescent protein (GFP) labelled siRNA. This knockdown method in dinoflagellates was initially challenging due to the large copy number of target DNA and permanently-condensed chromatin, but recent research has shown that there is a strong possibility that knockdown procedures can be successful.

A system for introducing antisense morpholinos into dinoflagellate cells without the use of PEG to warp the cell wall, or the use of RNAi was used in this study. Here, a novel delivery peptide was used to deliver morpholinos via an endocytosis-mediated process that avoids damaging the plasma membrane of the cell (Summerton 2005). This knockdown process can be done completely by immersion when peptide and antisense morpholino are added directly to the culture to stimulate endocytosis and morpholino uptake. To test this system on dinoflagellates, we used the manufacturer recommended concentrations of delivery-peptide and antisense-morpholino on a dense culture of *A. carterae*. Our preliminary data has shown that unlike with PEG addition, the delivery peptide does not cause the cell population to drastically decrease.

To test the effect of the morpholino addition on the dinoflagellate culture, antisense morpholinos were designed to target eukaryotic initiation factor 4E-1 (eIF4E-1). The eIF4E-1 is considered to be the main dinoflagellate translation factor (Batool *et al.* 2019; Jagus *et al.* 2012; Jones *et al.* 2015; Roy *et al.* 2018). Core dinoflagellate eIF4E translation factors are divided into 3 clades (1, 2, and 3), along with 3 subclades within each (a, b, c); with a total of 9 members. Our previous work has shown that these eIF4E family members diverge at critical amino acids, suggesting the family members are functionally distinct (Jones *et al.* 2015; Li & Wang 2005). Of these 3 major clades, eIF4E-1 stands out as the most duplicated, and with the lowest number of substitutions. Based on the expression levels of the subclades, our lab has theorized that eIF4E-1a is likely the primary translation initiation factor. We found that the use of antisense morpholinos can decrease the overall expression of this translation factor.

Methods

Cell Culturing

Amphidinium carterae (Hulbert) strain CCMP1314 was grown in ESAW artificial marine media with a salinity of 20 supplemented with f/2 nutrients without silicates at 25°C (Berges *et al.* 2001). The medium was buffered with 1mM HEPES (pH 8.0). Since bacterized cultures have shown to affect analyses of translation rate in *A.carterae*, the cultures were maintained with an antibiotic solution of kanamycin (50 µg/mL), carbenicillin (100 µg/mL), and spectinomycin (50 µg/mL) (Liu *et al.* 2017). The cultures were grown under 100 µmol/m^2 s^{-} light. Delivery of the morpholinos (described below) requires constant swirling to keep the reagents in solution, therefore the cultures were also placed on an orbital shaker at 60 rpm (Summerton 2005). The cultures were allowed to acclimate to the swirling for a week before knockdown reagents were added during log phase.

Morpholino Customization and Delivery

The sequence of the initiation factor eIF4E-1a was previously determined by our lab [GenBank: AMA65227.1] (Jones *et al.* 2015). Morpholino antisense oligonucleotides (MOs) are nucleic acid analogues in which DNA bases are bound to a non-charged backbone (morpholine rings linked by phosphorodiamidate bonds) (Todaro *et al.* 2021). For our purposes, a translation-blocking MO was created that covered the eIF4E-1a translational start site (5'-TCATTGAAGCTCAAACAAGCCATTG-3'). Specificity for the intended target sites was verified by BLAST analysis against the *A. carterae* transcriptome (Bachvaroff *et al.* 2014). The MOs were purchased from GeneTools (Philomath, OR, USA) and modified with a red-emitting fluorescent 3' Lissamine addition and then used at a concentration of 1 µM and 10 µM. Standard control MOs with the Lissamine addition were ordered as well (5'-CCTCTTACCTCAGTTACAATTTATA-3').

The MOs were delivered using Endo-Porter reagent (GeneTools, Philomath, OR) at a concentration of 4μM. Cultures of *A.carterae* seeded in 12-well plates were treated with either Endo-Porter, MO, or both. Three biological replicates of each treatment were performed.

Cell Counts and Fluorescence Quantification

Following the addition of 4 μM Endo-Porter and 1 μM or 10 μM MO, the viability of cultures of *A. carterae* were observed over a 96-hour period by measuring autofluorescence (640 nm) via flow cytometry. Measurements of the cultures within the first hour of treatment were labelled as Hour 0. Cell counts for each condition were determined on a BD C6 Accuri Flow Cytometer (BD Biosciences, San Jose, CA), equipped with laser excitation at 488 and 640 nm and emission at 533/30, 585/40, and > 670 nm. The FSC-A and fluorescence channels were used to select for live cells; from this selection, cells with Lissamine emission signals were detected (585 nm). Cells were grouped into low and high Lissamine fluorescence, with high fluorescence showing an intensity minimum of 10^4 relative fluorescence units (RFUs).

Cell Imaging

Images of *A. carterae* cells with and without MO treatment were taken on a STELLARIS confocal microscope (Leica Microsystems, USA), equipped with 405, 552 and 638 nm lasers, and PMT and HyD detectors collecting emission within 590-600 nm and 680-720 nm, respectively.

Quantification of Protein Expression

Protein expression was quantified by Western blotting. For the initial Western blot analyses at 1 μM morpholino concentration, the cell density of the cultures was quantified via flow-cytometry to create equal volume pellets containing ~75,000 cells for each condition 48 hours post-treatment. Cell pellets were collected midday and prepared for electrophoresis in 3× sample buffer (Blue Loading Buffer Pack, New England Biolabs, Ipswich, MA), heated to 96°C for 10

min and centrifuged for 2 min at 10,000 *g*; from the 15 µl total, 10 µl of each extract was electrophoresed at 165 V on a NuPAGE 4 to 12% Bis-Tris 1.5 mm Mini Protein Gel until the dye front reached the bottom of the gel. The gel was transferred to a membrane with the Trans Blot Turbo Transfer system. Protein loading and relative expression levels were verified by probing the same blot with anti-eIF4E-1a mouse monoclonal and HRP-conjugated anti-mouse IgG. The labeled bands, as well as band intensities, were detected with ImageLab software.

For the subsequent experiment testing the higher 10 µM concentration of morpholinos, equal quantities of whole-cell lysates containing 100,000 cells were prepared from each of the triplicate sample cultures at 48 and 96 hours. Samples were once again split equally over two NuPAGE 4-12% Bis-Tris Gels and run for 50 min at 165V. Both gels were transferred to a membrane with the Trans Blot Turbo Transfer system. To control for cell loading error, total protein was detected on one membrane using No-Stain Protein labelling Reagent (Invitrogen, Waltham, MA). The total protein was summed within each lane using ImageJ. Protein loading and relative expression levels were again verified by probing the second blot with anti-eIF4E-1a mouse monoclonal and HRP-conjugated anti-mouse IgG. The labeled bands were detected with ImageLab software, as well as band intensities. The relative production of eIF4E-1a was analyzed by dividing the eIF4E-1a volume by total protein. All conditions were performed in triplicate. Pairwise sample comparisons within timepoints were analyzed in R-Studio with t-tests using pooled standard deviations. A p-value of <0.05 was considered statistically significant.

Results

Cell Viability

Immediately post-treatment at T0, the culture growth was suspended, but recovery was observed at 24 hours (Figure viii). The cultures containing the antisense morpholino also showed

significant differences in growth compared to the control immediately after treatment addition and at 24 and 48 hours.

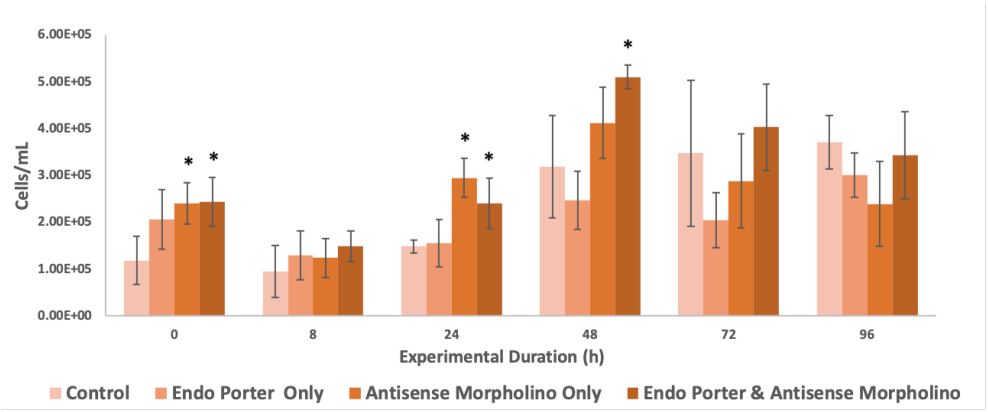


Figure viii. Flow cytometer population counts from Amphidinium carterae cultures (N=3). Conditions included the control, Endo-Porter only, antisense morpholino only, and the Endo-Porter and antisense morpholino combined. Statistical analysis was done using t-tests with pooled standard deviations.

** Significantly different from 'Control' based on a p-value of < 0.05.*

Uptake of the Morpholino

The intensity of the Lissamine signal within cells was measured by flow cytometry (Figure ix). Peak Lissamine fluorescence was observed at 48 hours, with about 13% of the population up-taking a high amount of fluorescently-tagged morpholinos and waned after this timepoint. The mean Lissamine fluorescence per cell with the Endo-Porter delivery system plus the MO was over 50X greater than that of the control, and the median was over 1.5X greater, showing significant uptake of the MO in *A. carterae*, as well as a large positively skewed distribution of uptake efficiency (Supp. Table v). Cells were also able to uptake morpholino without Endo Porter, but to a significantly lower degree, with no correlation to experimental duration.

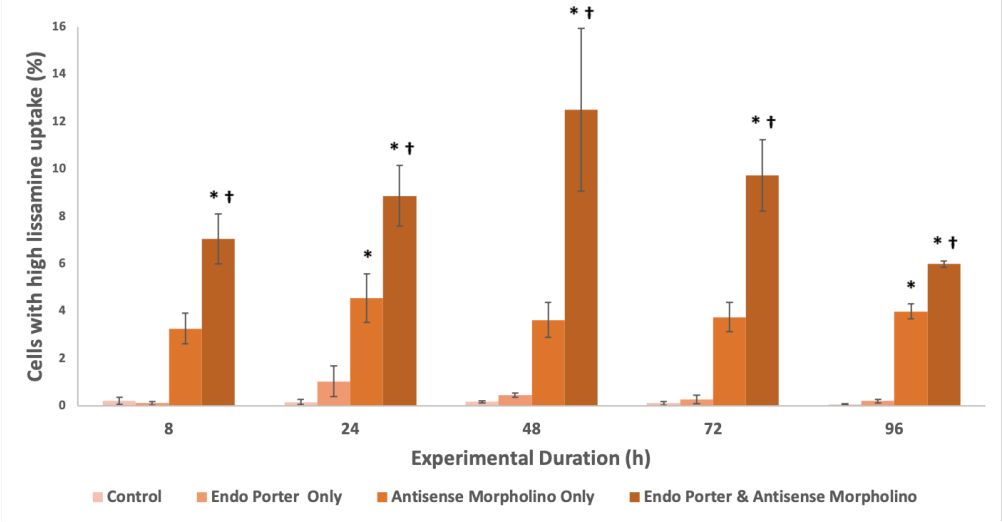


*Figure ix. Percent of A. carterae populations with high Lissamine uptake after 48 hours (cutoff 10^4 RFUs). Conditions included the control, Endo Porter only, antisense morpholino only, and the Endo Porter and antisense morpholino combined (N=3). Statistical analysis was done using t-tests with pooled standard deviations. * Significantly different from 'Control' based on a p-value of < 0.05. † Significant difference between "Antisense Morpholino Only" and "Endo Porter & Antisense Morpholino" based on a p-value of < 0.05.*

Images of cells were also captured by confocal microscopy (Figure x). Lissamine fluorescence was detected within the range of 590-600 nm (peak 593nm), and autofluorescence was measured between 700-720 nm. Confocal images display a diffuse pseudo-blue coloring within the cytoplasm of the dinoflagellate cells, as well as in a large area near the nucleus, indicating both diffuse and localized morpholino presence (Figure x).

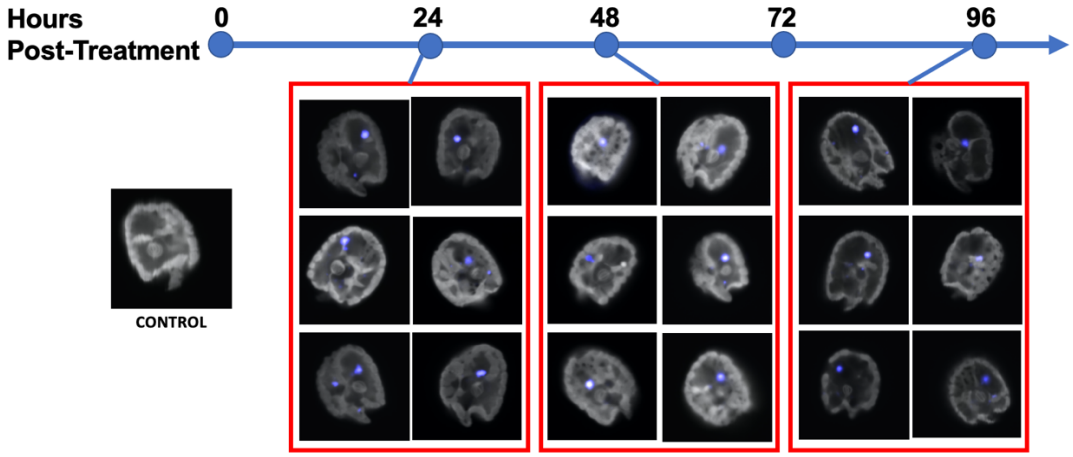


Figure x. Confocal images of Amphidinium carterae. A control cell with no antisense morpholino introduced is on the far left. Red boxes encompass images of cells from a culture with Endo Porter and antisense-morpholino, fluorescently tagged with Lissamine (593 nm, pseudo-blue) at 24-, 48- and 96-hours post-treatment. The grey areas are the auto-fluorescence emitted within 720-759 nm.

Initial Western Blot Analysis

From the initial Western Blot analysis, the expression of eIF4E-1a was significantly decreased ~30% compared to the control in cultures containing both the 1 µM morpholino and Endo Porter (Figure xi). Cultures with Endo Porter showed a decrease of 11% in eIF4E-1a expression over cultures without Endo Porter, although the difference was not significant.

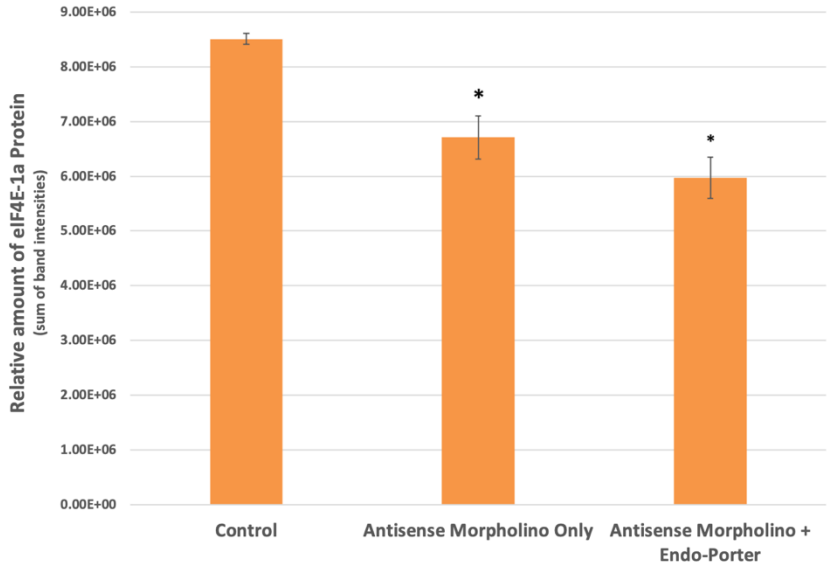


Figure xi. Western blot analyses for eIF4E-1a concentrations within control and treated cells at 48 hours post-MO [1μM] addition (N=3). Protein loading and relative expression levels was verified by probing with anti-eIF4E-1a mouse monoclonal and HRP-conjugated anti-mouse IgG. eIF4E-1a Protein area volumes were reduced by 30% after introduction of custom translation blocking morpholino and Endo Porter after 48 hours. Statistical analysis was done using t-tests with pooled standard deviations.

** Significantly different from "Control" based on a p-value of < 0.05.*

No significant difference found between "Antisense Morpholino Only" and "Antisense Morpholino + Endo Porter" based on a p-value of < 0.05.

Increase in Morpholino Concentration

Once the concentration of morpholinos was increased from 1 μM to 10 μM, the custom eIF4E-1a target morpholino and Endo Porter decreased eIF4E-1a expression at 48 hours compared to the control of about 42% (Figure xii). This was a greater decrease from the 30% found when using only 1μM of MO, constituting a total 11.8% decrease with the increased MO concentration (Table ii). The cultures containing eIF4E-1a-target morpholino only also showed reduced target-

protein production compared to the control of 21%. When compared to the culture containing only the standard non-target morpholino and Endo Porter, both of the cultures containing the eIF4E-1a-target morpholino with and without the Endo Porter appeared to exhibit a significant decrease of 47% and 29%, respectively (Figure xii). No significant difference was found between the cultures containing eIF4E-1a-target morpholino with or without the Endo Porter, although the cultures with the Endo Porter showed an average decrease of 25% compared to the cultures without.

Table ii. *Volume of eIF4E-1a produced compared to the control. Volumes of eIF4E-1a were quantified via western blot and compared to the density of a whole protein stain per sample (N=3). Percent of eIF4E-1a for each sample were compared to the control sample to describe relative eIF4E-1a production. Statistical analysis was done using t-tests with pooled standard deviations.*

Morpholino concentration	**1uM**	**10uM**	
	Relative eIF4E-1a production		**Percent change**
Antisense Morpholino Only	78.8%*	78.3%	0.5%
Antisense Morpholino + Endo Porter	70.1%*	58.3%*	11.8%

***** *Significant difference when compared to the control based on a p-value of* < 0.05.

Interestingly, we also found that expression levels became relatively similar after 96 hours, showing the temporary effects of the morpholino (Supp. Figure ii).

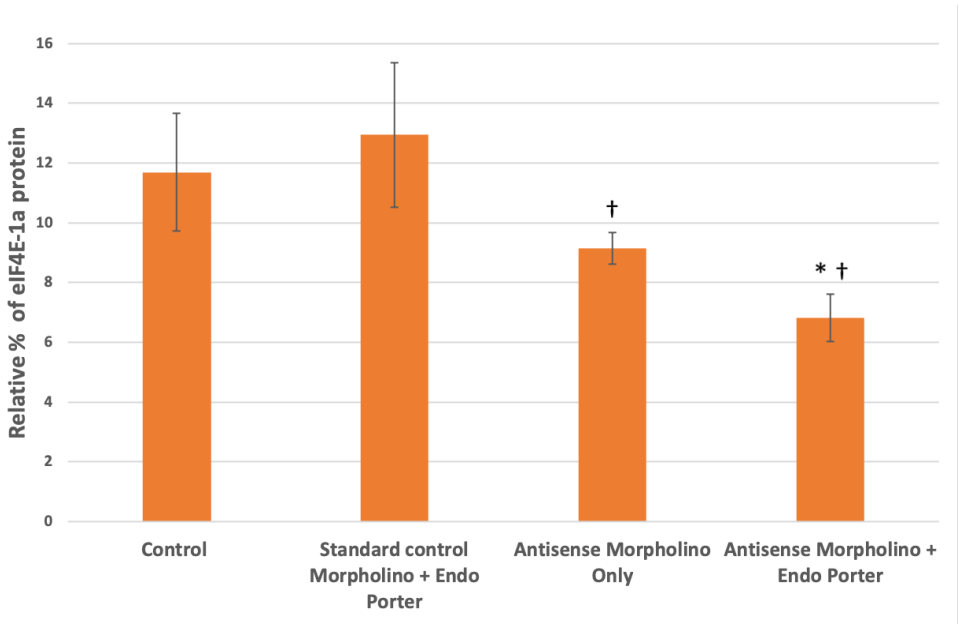


Figure xii. Western blot analyses for eIF4E-1a concentrations within control and treated cells at 48 post-MO [10 µM] addition (N=3). Protein loading and relative expression levels was verified by probing with anti-eIF4E-1a mouse monoclonal and HRP-conjugated anti-mouse IgG and compared to total protein volumes. Relative eIF4E-1a levels are lower in Amphidinium population after 48 hours of being subjected to custom translation-blocking morpholinos. Statistical analysis was done using t-tests with pooled standard deviations.

** Significantly different from 'Control' based on a p-value of < 0.05.*

† Significantly different from "Standard Morpholino + Endo Porter" based on a p-value of < 0.05.

No significant difference found between "Antisense Morpholino Only" and "Antisense Morpholino + Endo Porter" based on a p-value of < 0.05.

Discussion

Here a novel delivery peptide technology has allowed for successful introduction of custom antisense morpholinos into *A. carterae* cells without significantly decreasing viability. This method delivers substances into the cytosol of cells by an endocytosis-mediated process that avoids damaging the plasma membrane of the cell (Mellert *et al.* 2012; Summerton 2005). With this new system, exploration into the key players of different metabolic pathways may be closer than expected for dinoflagellates.

The above data was collected using the manufacturer recommended concentrations of reagents, but as we look further into these systems, we hope to optimize the outcome by adjusting protocol values. Currently, with 1 μM of MO added, we have been able to successfully introduce a high concentration of the MO into approximately 13% of the population (Figure ix), with a significant decrease in eIF4E-1a protein production. Dose-dependent effects of the MOs have also been observed, with an increase in MO concentration from 1 μM to 10 μM resulting in a decrease in eIF4E-1a from 30% to 42%; an 11.83% decrease in total (Table i). Optimistically we would like to increase this percentage to a level where protein production is functionally suspended or be able to sort cells from the population based on their MO uptake.

Interestingly, western blot analyses for both the initial 1 μM and subsequent 10 μM MO experiment showed no statistically significant difference between the cultures with and without Endo-Porter, although the cultures with the Endo-Porter showed consistently lower average eIF4E-1a protein produced (Figures xi and xii). This would indicate that even without a delivery peptide, *A. carterae* cells are able to uptake the MO. Reasons for MO uptake without a delivery peptide are still unclear. Over a million identified peptide sequences within various dinoflagellates are still of unknown function and origin, making their evolutionary history ambiguous (Stephens *et al.* 2018). Dinoflagellates are thought to have undergone multiple

organellogensis events, where the genome of endosymbiotic algae becomes a plastid and/or genes from the endosymbiont are transferred to the nucleus (Sarai *et al.* 2020). The evolutionary nature of dinoflagellates to accept foreign genes appears to be high (Morden & Sherwood 2002). One theory for why the *A. carterae* cells took up the unaided MO could be the possibility that dinoflagellate systems are more open to horizontal gene transfer (HGT) than previously imagined (Gagat *et al.* 2017). Recent studies have shown that genomes within dinoflagellates may be more open to foreign contributions from both bacteria and eukaryotes compared to other organisms (Wisecaver *et al.* 2013). Further research needs to be conducted to follow how foreign genetic material is received by dinoflagellates, as well as how transcripts are processed.

A wide distribution of MO uptake between cells may account for the discrepancy of measured eIF4E-1a protein production compared to the cellular MO uptake (Supp. Table v). Although we suggest about 13% of the cells have a high uptake after addition of 1uM of MO, the eIF4E-1a production was decreased by about 30% according to the western results (Figure xi). We assume that there is a range of efficiency in uptake within the population, so some cells may be producing little or no eIF4E-1a, while others may be producing their average amounts. Once again, changes to the concentrations used or the use of a cell-sorter may be necessary to observe higher gene knockdown efficiency.

The MO localizes at distinct sites within the *A. carterae* cells, usually in a large area by the nucleus (Figure x). In previous studies, Endo-Porter sometimes results in unsuccessful endosomal acidification, and therefore no release of the MO into the cytoplasm from their vesicle, but these localization points are usually known to appear as punctate fluorescence throughout the cell (Summerton 2005). The singular, large area of localization may signify MO aggregation in the nucleolus or an RNA-granule (Gavelis *et al.* 2019; Gornik *et al.* 2019; Lee *et al.* 2020; Sarai *et al.* 2020; Soyer-Gobillard & Geraud 1992). Dinoflagellates are known to have very unusual nuclei,

specially named the “dinokaryon.” Among other peculiar features are recently discovered “nuclear tunnels” which extend from the nuclear envelope of *Polykrikos kofoidii*, specifically during mitosis (Gavelis *et al.* 2019). These nuclear envelope tunnels are also connected with a membranous structure throughout the nucleus known as the “nuclear net”. The discovery of these structures adds a new level of complexity to the dinoflagellate nuclear membrane and may allude to more complex processing for transcription and translation. The unusual nuclear envelope tunnels along with our MO localization could very well be connected. Since dinoflagellates are known to regulate gene expression at the translational level, this cellular organization of mRNA may be a crucial step for gene expression, which needs to be analyzed further.

In theory, this method of gene knockdown expands research into the biosynthetic pathways within dinoflagellate. Currently, this research is sorely lacking, largely due to their complex genomes and unusual cell biology. Translation regulation is now understood as a crucial step in gene expression, far beyond that of transcriptional control (Jones *et al.* 2015; Roy *et al.* 2018). In addition, further understanding of the components of the translation machinery is required to understand the expression of specific genes, which make up the dinoflagellate “translational toolkit”. As we work to optimize this procedure, other pathways could be targeted as well in other species of dinoflagellates.

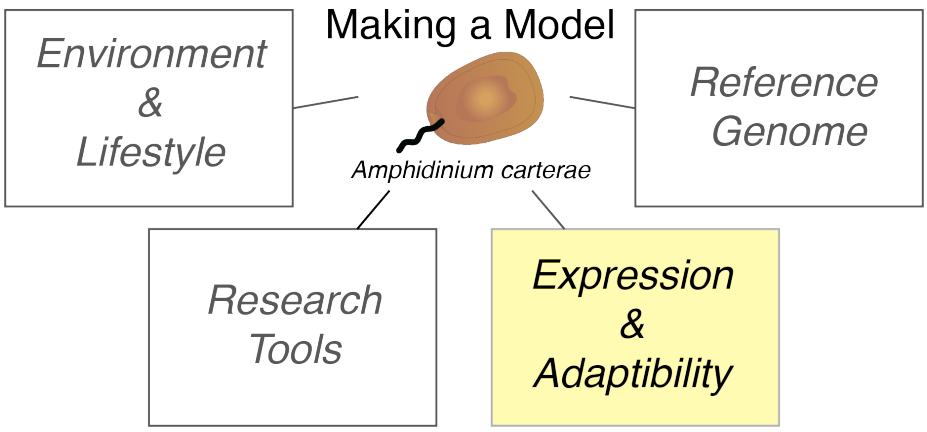
One interesting area would be to target a transcript responsible for toxin production. The function of toxins produced by dinoflagellates has been theorized but is still unknown. Theories include allelopathy, prey-capture, and as a defense (Sheng *et al.* 2010). One way to discern the function of the toxins would be to knockdown their expression to see how this affects feeding and swimming behavior. Mixotrophic characteristics of dinoflagellate species known to produce toxins could be monitored through photosynthesis and respiration rates, as well as swimming

behavior via digital holographic microscopy. This data could produce more evidence to the intended functionality of dinoflagellate toxins.

Conclusions

This primary study provides proof-of-principle for the possibility to specifically down-regulate gene expression in dinoflagellates using antisense morpholinos and a novel delivery system. Additional work is necessary to validate and optimize these findings, and to extend them to other biosynthetic pathways. Further work will also be required to investigate the effects of gene knockdown on the various eIF4E family members in order to possibly unveil a translation toolkit used by dinoflagellates to regulate gene expression. Studies involving toxin biosynthesis pathways would also benefit from successful knockdowns in order to identify functionality and key synthesis steps. This method of genetic knockdown increases the research tools that are available for dinoflagellates, and its success in *A. carterae* in the current chapter adds to the use of this dinoflagellate as a potential model.

Chapter 4: Dinoflagellate Gene Expression



Introduction to Dinoflagellate Genomes

Understanding gene expression, diel cycling, and cell cycle regulation is essential when establishing a model organism because these processes define how the organism functions, adapts, and responds to its environment. Expression patterns reveal how genes are regulated, and which pathways are active under different conditions, offering insight into the organism's core biology. Diel cycling, the rhythmic changes tied to light–dark cycles, is particularly critical in photosynthetic and aquatic organisms, as it governs metabolism, energy production, and ecological interactions across a day (Leighfield & Van Dolah 2001; Lin 2024). Similarly, cell cycle dynamics determine growth, division, and responses to environmental cues, making them fundamental to experimental reproducibility and comparative studies. By characterizing these layers of regulation, researchers can better predict behavior, identify conserved versus unique traits, and use the organism more effectively as a representative system for broader biological questions.

In dinoflagellates, examining expression, diel cycling, and cell cycle control is especially important because their biology departs significantly from classical model systems. These organisms exhibit unusual genomic and regulatory architectures, such as permanently condensed

chromosomes, widespread gene duplication, and constitutive transcription with much of the control occurring post-transcriptionally (Lin 2024). Their physiology is tightly coupled to diel rhythms, with processes like photosynthesis, bioluminescence, and toxin production often oscillating across light–dark cycles. Likewise, their cell cycles are closely synchronized to environmental cues, including diel light patterns, making them an excellent system to study how external rhythms integrate with internal regulation. By characterizing these unique features, we can not only advance a dinoflagellate as tractable models but also gain insight into alternative strategies of gene regulation and adaptation that expand our understanding of eukaryotic diversity.

Methods

Cell Culturing

Cultures of *Amphidinium carterae* (Hulbert 1314) were created in 200mL of ESAW artificial marine media with a salinity of 20 supplemented with f/2 nutrients without silicates at a starting concentration of 10K cells mL^{-1} and allowed to acclimate for 5 days in 14:10 h light:dark period in 100 μmole photons m^{-3} sec^{-1} at 25ºC (Berges *et al.* 2001). Cultures were then kept in the dark for 48 hours to retard division and synchronize the population (shown by nuclei staining below) before being reintroduced to the previous light phases for sampling over 24 hours (Leighfield & Van Dolah 2001).

Two diel experiments were conducted with some overlapping sample processing as well as unique analyses per trial (Figure xiii-A). In the first experiment, three flasks were synchronized, and samples were taken at 0, 4, 8, 12, 16, 20, and 24 hours starting at the beginning of the light phase for nuclei staining, puromycin-incorporation, and protein abundance, with the dark-phase beginning at 14 hours. Samples for Nanopore direct RNA sequencing were also taken at 6, 12, and 20 hours. In the second experiment, 12 flasks were created for Ribo-seq and RNA-seq

sampling only and sampled at 6, 12, 16, and 20 hours. Three more flasks were sampled at 0, 3, 6, 9, 12, 16, 20, and 24 hours for nuclei staining and at 6, 12, 16, and 20 hours for puromycin-incorporation (Figure xiii-B). Cells were counted using a Scepter™ 2.0 Handheld Automated Cell Counter using a 40 µm sensor. For naming purposes, the three timepoints that included RNA sampling will be named according to their timepoint in the cell cycle and their light-phase: "Growth - Midday", "Synthesis - Evening", and "Division - Midnight" (Figure xiii-C).

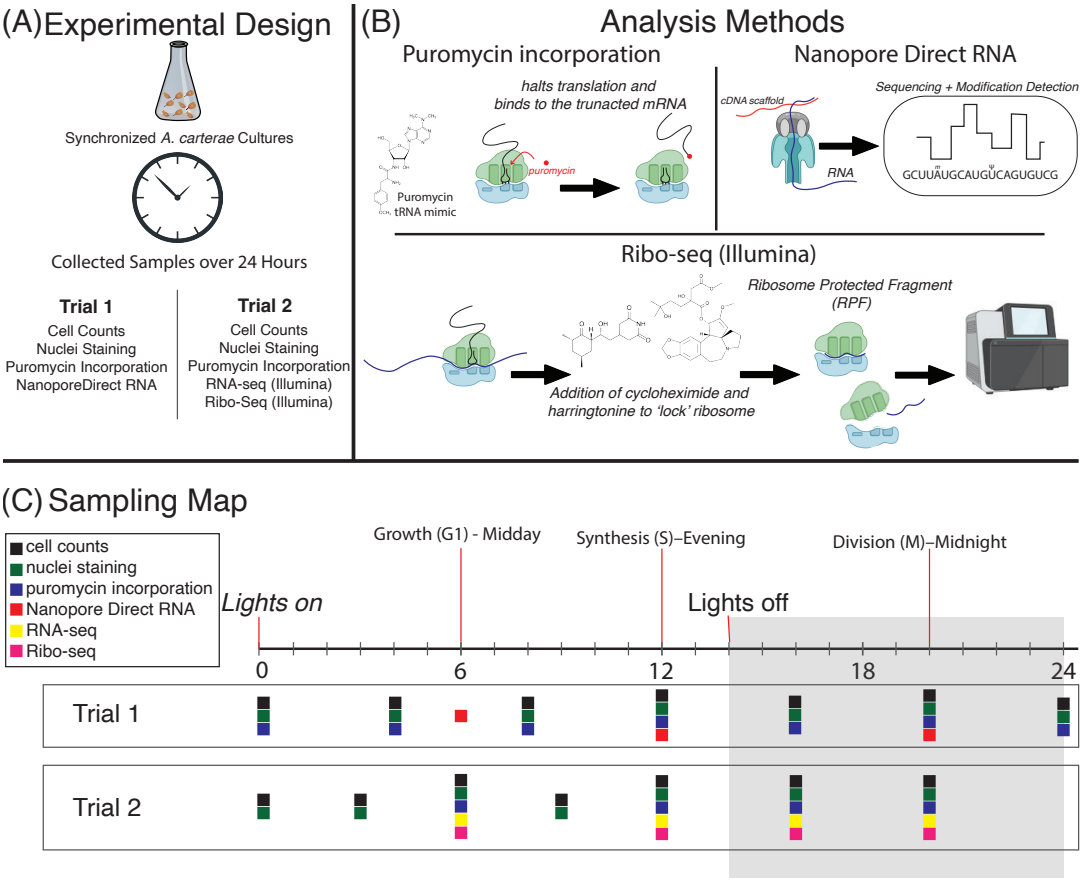


Figure xiii. Experimental Design for Expression Measurements. (A) Broad overview of 24-hour length sampling. (B) In-depth description of complex sample processing including puromycin incorporation (as a proxy for translation rates), Nanopore Direct RNA sequencing, and Ribosome profiling (Ribo-seq). (C) Timepoint map of sample processing.

Nuclei Staining

To measure DNA content per cell, triplicate aliquots of 1mL of culture were spun down at 2000 x g for 10 minutes and the media was discarded. The cell pellet was resuspended in -20ºC methanol and stored at -20ºC. Once all samples were collected over the diel cycle, the samples were spun at 2000 x g for 10 minutes, the methanol was discarded, and the cell pellet was resuspended in 500 ml of PI solution (base intercalating dye; 10 mg mL^{-1} Propidium Iodide in phosphate buffered saline containing 40 units ml^{-1} RNase and 0.5% Tween 20) (Leighfield & Van Dolah 2001). The nuclei-stained sample was then analyzed via flow cytometry with a BD Accuri™ C6 Flow Cytometer using the FL2 channel 488nm laser.

Translation measurements: SUnSET method

Puromycin incorporation was used as a proxy measurement for translation rates for one flask in a method known as surface sensing of translation (SUnSET) (Goodman & Hornberger 2013) (Figure xiii-B). Puromycin is an aminonucleoside antibiotic produced by *Streptomyces alboniger* that is a structural analog to tRNA. Within the cell, puromycin incorporates into growing peptide chains and halts translation, resulting in a released peptide ending with a puromycin molecule. Abundance of these neosynthesized puromycin-containing peptides directly reflects translation rates when measuring with an antibody to puromycin. At each sampling point, aliquots of the culture were incubated with 1 mM puromycin for exactly 45 minutes to allow for detectable incorporation. Aliquots were then spun down and stored in SDS-PAGE buffer at -20ºC before immunoblot analysis. Cell pellets were prepared for electrophoresis in 3× sample buffer (Blue Loading Buffer Pack, New England Biolabs, Ipswich, MA), heated to 96°C for 10 min and centrifuged for 2 min at 10,000 *g*; 10 µl of each extract was electrophoresed at 165 V on a NuPAGE 4 to 12% Bis-Tris 1.5 mm Mini Protein Gel until the dye front reached the bottom of the gel. The gel was transferred to a PVDF membrane with the Trans Blot Turbo Transfer system. Detection of incorporated puromycin was measured in triplicate on an immunoblot using a

monoclonal mouse anti-puromycin antibody (Sigma Aldrich, MO, U.S.A) and a secondary goat anti-mouse HRP monoclonal antibody (Invitrogen, California, U.S.A). Normalization was done using No-Stain™ Protein Labeling Reagent (Invitrogen, California, U.S.A) to visualize the total protein in each lane.

RNA Processing

For the Nanopore Direct RNA sequencing, 30 mL of culture was centrifuged at 10000 x g for 10 min to pellet cells (Figure xiii-B). The media was discarded, and the pellet was resuspended in 1 mL of Trizol and stored at -80°C. After thawing, 0.2 mL of chloroform was added to the sample and inverted to mix. This was centrifuged at 12000 x g for 15 minutes at 4°C before the aqueous layer was transferred to a new tube. Then 0.5 mL of isopropanol and 5 ug of glycogen were added and inverted to mix, followed by a 10-minute incubation at 4°C. The sample was then mixed at 12000 x g for 10 minutes at 4°C and the supernatant was removed. The RNA pellet was resuspended in 1mL of 75% ethanol and vortexed briefly before being centrifuged at 7500 x g for 5 minutes at 4°C. The supernatant was removed, and the pellet was allowed to dry for 10 minutes at room temperature. The pellet was then resuspended in 20uL of RNase-free water and incubated for 10 minutes at 55°C. DNA contamination was removed by 1.8 U adding DNase I (Thermo Scientific) and 40 U of RNaseOUT Recombinant Ribonuclease Inhibitor (Invitrogen) and incubating at 37°C for 30 minutes, followed by an incubation at 65°C for 10 minutes. Samples integrity was assessed via Bioanalyzer at this point (Supp. Figure iii). Since dinoflagellates-derived RNA commonly has chloroplast rRNA contamination as well, RIN numbers close to 6 were advanced to the next stage, with more emphasis being placed on manually checking for the presence of the 18S/28S rRNA peaks (Akbar *et al.* 2018b). The sample was then stored at -80°C. A NEBNext Poly(A) mRNA Magnetic Isolation Module kit was used to remove rRNA and isolate polyadenylated mRNA from the thawed samples. Briefly, this included 4 washes of the mRNA sample with oligo dT beads with RNA Binding Buffer, followed by 2 washes with a wash

buffer, and a final elution in Tris buffer. From here, Direct RNA libraries were prepared for each sample with Oxford Nanopore's SQK-RNA004 kit and loaded onto a FLO-MIN004RA flow cell.

Short-read Illumina RNA-seq data was produced by collecting over 1 million cells in 20 mL of culture and centrifuging for 5 min at 2000 x g. The supernatant was then discarded, and the pellet was resuspended in 1mL of Trizol and flash-frozen in liquid nitrogen before being stored at -80°C. These samples were then sent to CD Genomics for sequencing.

Ribosome profiling samples were collected in tandem with the Short-read RNA-seq samples, with over 8 million cells collected in 160 mL of culture (Figure xiii-B). The sample was then incubated with 2ug/mL harringtonine for 2 minutes before immediately being centrifuged at 2000 x g for 5 minutes. The supernatant was discarded, and the cells were resuspended in 200uL of 4C PBS with 2ug/mL harringtonine and 100ug/mL cycloheximide. The sample was then flash-frozen in liquid nitrogen and stored at -80C before being sent to CD Genomics for Ribosome Protected Fragment (RPF) isolation and sequencing.

RNA Analysis

Illumina RNA-seq and Ribo-Seq reads were trimmed for adapters, with average lengths of 150 and 30 bp, respectively, and filtered for quality (phred >30) with fastp (Table iii). A transcriptome was assembled using the RNA-seq reads with Trinity, and the transcripts were further collapsed into super-transcripts. Both RNA-seq and Ribo-seq datasets were mapped via bowtie2 using end-to-end mapping and no greater than 1 mismatch to the bacterial genomes described in Chapter 2 to remove bacterial contamination. The RNA-seq and Ribo-seq reads were then mapped via bowtie2 using end-to-end mapping and no greater than1 mismatch to the short-read trinity assembly of super-transcripts. Read mapping was quantified using featureCounts and finally, differential expression analysis was run using edgeR with replicates made *in silico* (Liao

et al. 2014; Robinson *et al.* 2010). Translation efficiency was calculated using a regression-based Analysis of Partial Variance (APV) using the anota2seq R package (Oertlin *et al.* 2022).

Nanopore RNA Direct Sequencing reads were processed using Dorado 0.8.3, and basecalled with the rna004_130bps_sup@v5.0.0 model with modification calling for N6-Methyladenosine (m^6A), 5-methylcytosine (m^5C), and pseudouridine (pseU). Reads were filtered for phred quality (>10) using chopper (De Coster & Rademakers 2023) (Table iii). Abundance of the transcripts was quantified by read mapping via minimap to the transcriptome, followed by quantification with salmon with a TPM cutoff of 1. The R package edgeR was then used for differential expression analysis with *in silico* triplicates. Correlation of modifications to expression was calculated using a Spearman test between modification fraction and expression, a Wilcoxon test on expression, and a Fisher test on the differentially expressed genes with their modification loss/gain. The combined read files were also mapped to the Judd *et al.* (2025) *A. carterae* reference genome using minimap2 (See Chapter 5) (Li 2018).

Table iii. RNA Reads	**Growth (G1) - Midday**	**Synthesis (S)–Evening**	**Division (M)–Midnight**
Nanopore Direct RNA (single)			
Initial Read Count	5,163,621	4,017,730	4,578053
Read Count Post-filtering	4,389,278	3,434,333	3,744,999
Total Nanopore Reads	**11,568,610**		
Illumina RNA-seq (paired)			
Initial Read Count	22,937,135	30,394,118	23,240,973
Read Count Post-filtering	22,706,696	30,218,145	23,062,121
Illumina Ribo-Seq (paired)			
Initial Read Count	185,879,754	249,608,011	57,410,424
Read Count Post-filtering	161,581,049	131,993,820	51,180,373

Results

Nuclei Staining

The population size within the cultures remained relatively constant over the day, with a doubling occurring during the dark phase, between 20 to 24 hours after the experiment start (3-7 hours into

the dark phase). Culture counts showed an average population increase of 43%, which remained relatively constant into the next day. Concurrent nuclei staining of these timepoints showed that less than 10% of the population had individuals with double DNA content up until 12 hours into the experiment when this rose to an average of 11.85% and peaked at 16 hours (2 hours into the dark phase) at 26.37% (Figure xiv-A). Then the percent of individuals with doubled DNA content decreased at 6 hours into the dark phase to 16.45% and returned to 6.07% at 24 hours into the experiment, just as the light phase was about to begin. The population doubling in the middle of the night along with the general increase in DNA content within the cells displays a relatively well-coordinated population uptick in DNA in preparation for cell division, with division occurring throughout the population between 16 and 20 hours.

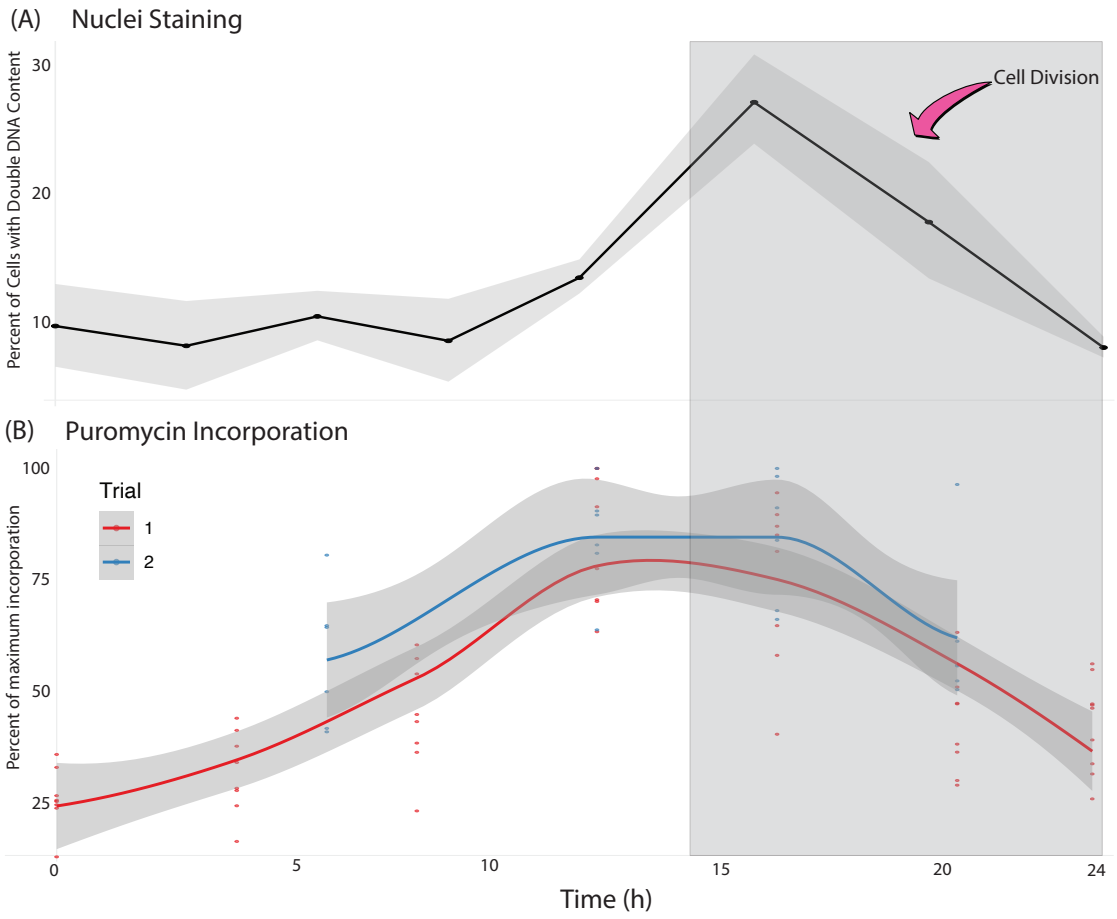


Figure xiv. DNA Replication & Translation over the Diel cycle. (A) Measurements represent percent of the cell population with doubled DNA content fluorescent signal based on Propidium Iodide staining. (B) Relative puromycin incorporation over the diel cycle, relative to the largest amount of incorporated puromycin detected during the day, which was set to 100%. The first round of sampling is seen in red, and the second round of sampling is seen in blue. Triplicates graphed in R with a local regression trendline.

Puromycin Incorporation

Using puromycin-incorporation as a proxy for translation rates, we demonstrated continued protein synthesis throughout the cell cycle with a broad peak in the evening during the Synthesis phase (Supp. Figure iv). With the maximum blot lane intensity per gel being designated as 100%, we observed a steady increase in puromycin-incorporation from 0 to 16 hours from 25.3 ± 6.87% to 85.1 ± 13.7% followed by a steady decrease during the dark phase to 24 hours, showing an increase in translation during the daytime Growth phase into the Synthesis phase, with a decrease in translation into the nighttime Division phase. Overall, the most puromycin-incorporation, representing the peak of protein production, occurred right before the end of the day (Figure xivB).

Short-Read Transcriptome Assembly

The transcriptome assembled from the RNA-seq reads initially had 108,640 transcripts which were compressed into 65,900 super-transcripts. These transcripts were then filtered (to remove any contamination due to multiplexing) based on identity using BLASTn's default parameters to an earlier *A. carterae* transcriptome created by Bachvaroff *et al* 2014 (Bachvaroff *et al.* 2014), leaving 53,434 transcripts shared between the datasets. The final transcriptome had an N50 of 2,146 with a range from 500 to 45,224 bases. Overall, the number of transcripts with a

transcripts-per-million (TPM) value >1 was 45,873 (85.8% of transcripts) with a maximum TPM of 10,238 and a TPM mean of 18.7.

Nanopore Direct RNA Expression

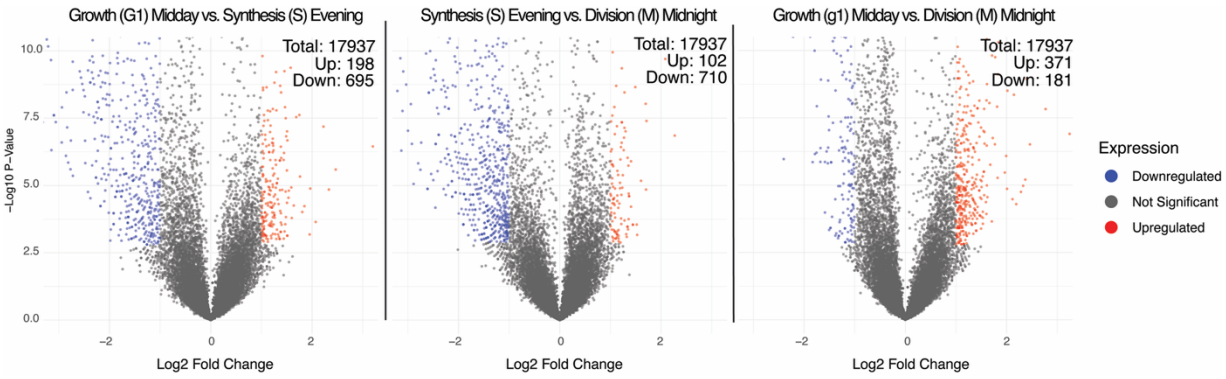


Figure xv. Nanopore Direct RNA differential expression across timepoints. Panel A: G1 versus Synthesis timepoints, Panel B: Synthesis versus Division timepoints, and Panel C: G1 versus Division timepoints.

The direct RNA sequencing showed only small overall number of genes that had narrow ranges of differential expression within the predicted transcripts**.** Of the 53,434 assembled super-transcripts, 17,937 had expression levels over 1 TPM for all 3 timepoints (Figure xv). Overall, there were 1,627 unique transcripts that displayed significant differential expression over the 3 timepoints with log fold change (logFC) >1 and p-value < 0.05. The logFC for each timepoint ranged from -2.4 to +3.2 for 5% of expressed genes, -3.8 to +2.3 for 4.5% of expressed genes, and -3.6 to +3.2 for 3.1% of expressed genes for Growth versus Synthesis, Synthesis versus Division, and Growth versus Division, respectively. From G1 to Synthesis, 1.1% and 4% of transcripts saw increased or decreased expression, respectively. From Synthesis to Division, 0.6% and 4.1% of transcripts saw increased or decreased expression, respectively. And lastly for overall shifts from G1 to Division, 2.1% and 1% of transcripts saw increased or decreased

expression, respectively. For each comparison, 16 to 159 transcripts displayed differential expression over logFC of 2, displaying conservative differential expression.

Modified Bases

Nanopore Direct RNA reads with m^6A, m^5C, and pseU modified bases were quantified over the diel cycle. Based on a random sampling of 100K transcripts from all 3 timepoints, modification ranged from 5.1, 3.9, to 3.9% of m^6A for the Growth, Synthesis and Division phases. The pseU modification ranged from 5.5, 5.0, and 4.8% for Growth, Synthesis and Division phases. Lastly, the m^5C modifications ranged from 4.1, 4.9, to 5.1% for Growth, Synthesis and Division phases.

The RNA base modifications were also compared to transcript expression (TPM) showing a global trend of transcripts with at least one modification displaying higher abundance (Figure xvi-A). On average, transcripts with m^6A showed 0.12-0.25 X abundance increase, transcripts with m^5C showed 1.5-1.7 X higher abundance increase, and transcripts with pseU showed 1.5-2 X higher abundance increase. The different timepoints also displayed differential regulation of modified transcripts, with transcripts containing any modification being upregulated for Growth (Midday) to Synthesis (Evening) and also downregulated with the loss of the modification over the timepoints. For Synthesis (Evening) to Division (Midnight), gaining a modification between the timepoints was actually more associated with downregulation, and for overall shift from Growth (Midday) to Division (Midnight), gaining a modification was correlated with upregulation, but not statistically correlated with loss/downregulation. Overall, modification led to greater expression with some instances of RNA modifications being context dependent.

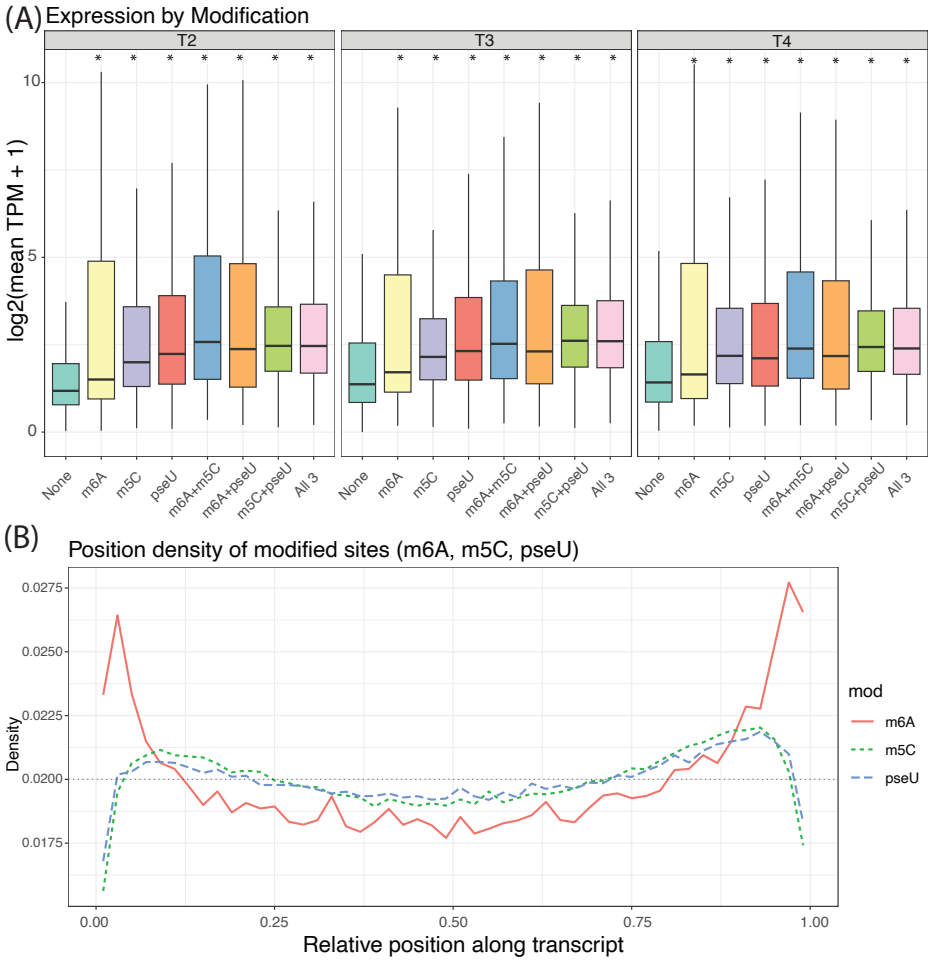


Figure xvi. Distribution of transcript modifications. (A) For each timepoint, expression distributions are shown for transcripts with no modification, m6A, m5C, pseU, and all combinations of modifications. means statistically different from the transcripts with no modifications. (B) distribution of modifications along the transcripts, normalized by length.*

The location of modification type varied over the transcripts. The m^6A modification appeared to be most dense at the ends of the transcripts with decreasing density towards the middle, whereas the m^5C and pseU modifications appeared to be relatively well distributed over the transcripts (Figure xvi-B). Orientation of the transcripts is generally from the 3' end where the sequencing adapter is ligated, but may not always be the same, so distinguishing the accumulation by 3' or 5' UTR is still unknown.

Polycistronic Transcription

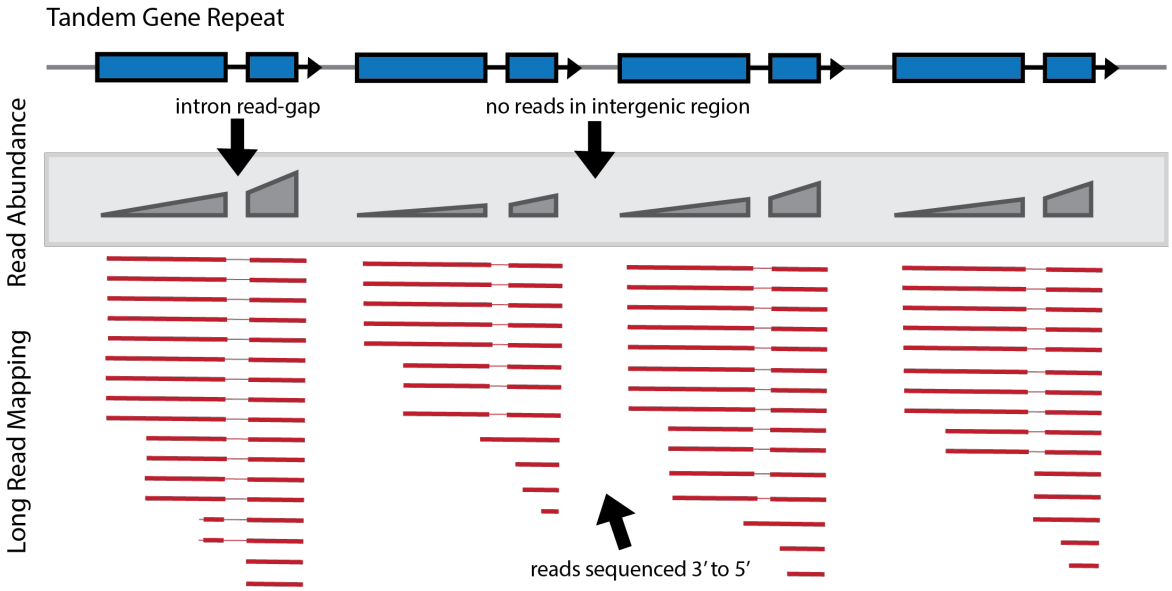


Figure xvii. Diagram of Nanopore Direct RNA read mapping to the A. carterae reference genome. The grey line represents the reference contig, the black lines represent tandem duplicate genes, and the blue boxes represent exons. The grey boxes represent read mapping abundance int that region and the red lines represent raw Nanopore Direct RNA read coverage, with thinner red lines showing an alignment gap.

Evidence of polycistronic transcription was assessed using mapping of Oxford Nanopore Direct RNA reads to a recent genome assembly for *A. carterae* (See Chapter 5). There are 68 identified tandem repeat gene families over 894 contigs, which are ideal for testing polycistronic transcription because they are densely packed with very short intergenic regions, highly expressed and are well-annotated. Overall, 789,950 Nanopore Direct RNA reads mapped to these tandem repeat genes, of which 237 mapped over an intergenic region (0.03%). None of the reads mapped completely through intronic regions of these gene groups (Figure xvii). The tandem repeat family with the largest amount of multi-copy mapping was the Major basic nuclear proteins (MBNP), which span over 2 contigs with 19 and 32 copies, respectively. The contig with

19 copies had 51,321 reads mapped with 83 mapping across a full intergenic region (0.16%) and the contig with 32 copies had 76,503 reads mapped with 32 mapping across the full intergenic region (0.04%). In the case of the MBNPs, there is some evidence that the gene copies follow a pattern of 2 copy types (see Chapter 5), which may explain why these regions could potentially be transcribed together, although separate transcription seems to be much more common. In the rare cases where a read mapped across an intergenic region, it never spans more than 2 copies of the genes.

Illumina RNA- and Ribo-seq

Differential Expression analysis differed between RNA-seq and Ribo-seq read sets due to the nature of the RNA being sampled in each method – either the full transcript pool or the actively translated transcripts, respectively. For the RNA-seq comparisons, the maximum LogFC ranged from -5.4 to 5.4, with 1,536 genes displaying an absolute LogFC in expression over 2 over the three comparisons and an average absolute LogFC of 2.0 ± 0.62 (Table iv).

The Ribo-seq data had a much higher range of differential expression, with maximum LogFC ranging from -9.1 to 7.7 involving 198 unique genes and an average absolute LogFC of 2.66 ± 2.36.

Table iv. Differentially Expressed Genes		**Growth vs Synthesis**	**Synthesis vs Division**	**Growth vs Division**
RNA-seq	Up Down	309 1714	683 377	1037 367
Ribo-seq	Up Down	83 55	75 66	51 73
TE	Up Down	68 45	58 51	42 64

Ribo-seq sequences aligned in two distinct groupings to the transcriptome, with 23.03%, 47.92% and 51.52% mapping across 25-30 bases and 76.22%, 46.56%, and 48.27 mapping across 30-40

bases for timepoints corresponding to Growth, Synthesis, and Division, respectively (Supp. Figure v).

Although both datasets for RNA-seq and Ribo-seq observed transcripts with statistically significant differential expression, the pools of transcripts seemed to vary more between the Ribo-seq data compared to the RNA-seq data. In total, the 3 timepoints shared 2% and 96% of the Ribo-seq and RNA-seq pools, respectively (Figure xviii-A and xviii-C). The pool of transcripts expressed in the RNA-seq data shared between only 2 timepoints ranged from 0-1%, while the transcripts expressed in the Ribo-seq data ranged from 0-7%. Lastly, we saw a trend of the RNA-seq pools for each timepoint lacking uniquely expressed transcripts, yet 73% of the expressed transcripts for Ribo-seq were expressed uniquely in the last timepoint corresponding to Division.

Although more transcripts were uniquely expressed in the last timepoint in the Ribo-seq data, many of these had very low expression, whereas the 701 transcripts uniquely expressed in the middle timepoint had a larger distribution of expression (Figure xviii-B). This difference was not reflected in the RNA-seq data, as transcripts unique to the individual timepoints had similar distributions and were all relatively lowly expressed (< 30 reads mapped) (Figure xviii-D).

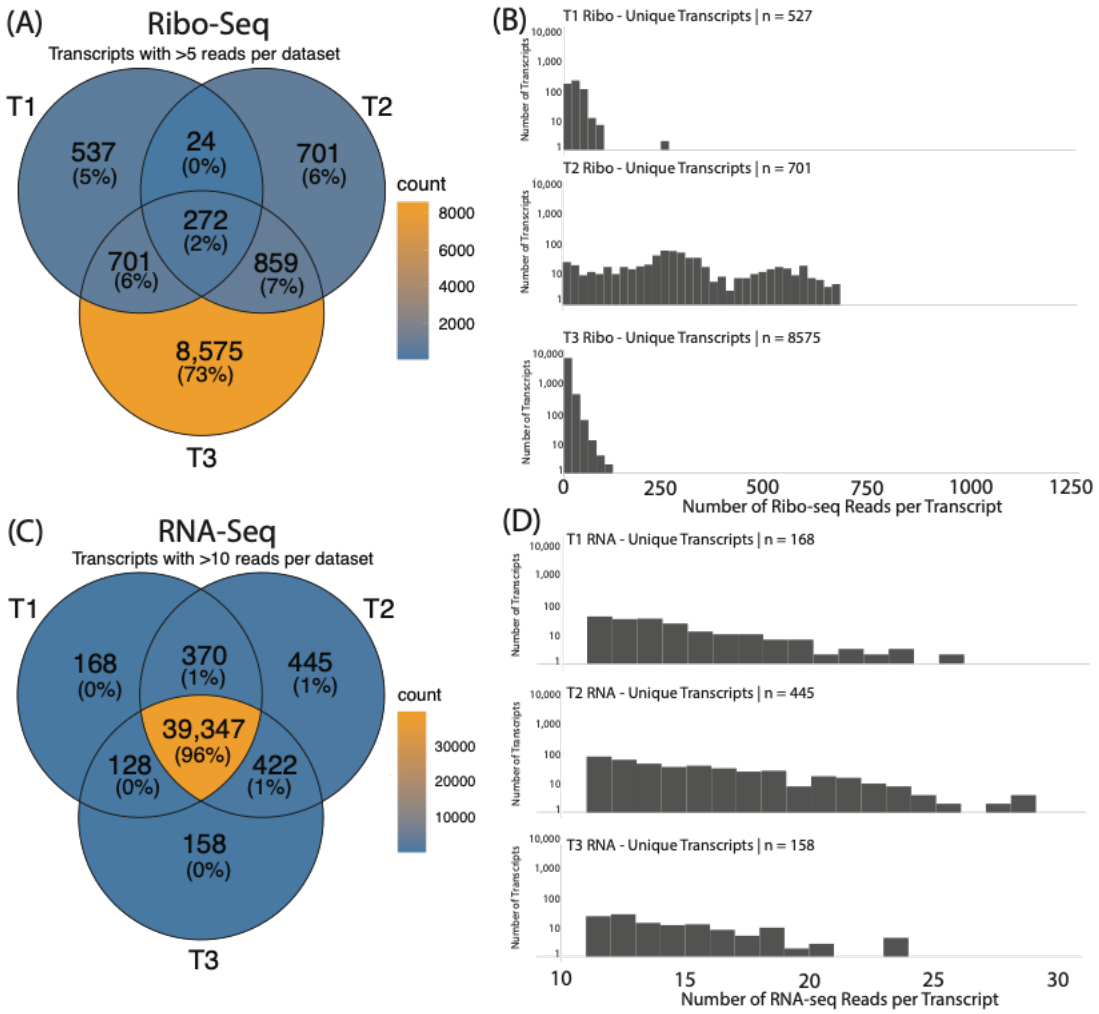


Figure xviii. *Shared transcript pools. (A) Venn Diagram of transcripts expressed in the Ribo-seq data across the 3 timepoints with shared and unique transcripts shown as numbers within each sector, (B) Distribution of read depth for the uniquely expressed transcripts in the Ribo-seq data for each timepoint, (C) Venn Diagram of transcripts expressed in the RNA-seq data across the 3 timepoints with shared and unique transcripts quantified, (C) Distribution of read depth for the uniquely expressed transcripts in the RNA-seq data for each timepoint.*

Fold-change distributions for the mapped RNA-seq data did not seem to vary globally, with the median absolute LogFC being 0.28, 0.22, and 0.29 for Growth versus Synthesis, Synthesis versus Division, and Growth versus Division, respectively (Figure xix). On the other hand, the

transcripts with Ribo-seq mapping had a median absolute log fold-change of 1.15, 1.24, 1.21 for Growth versus Synthesis, Synthesis versus Division, and Growth versus Division, respectively.

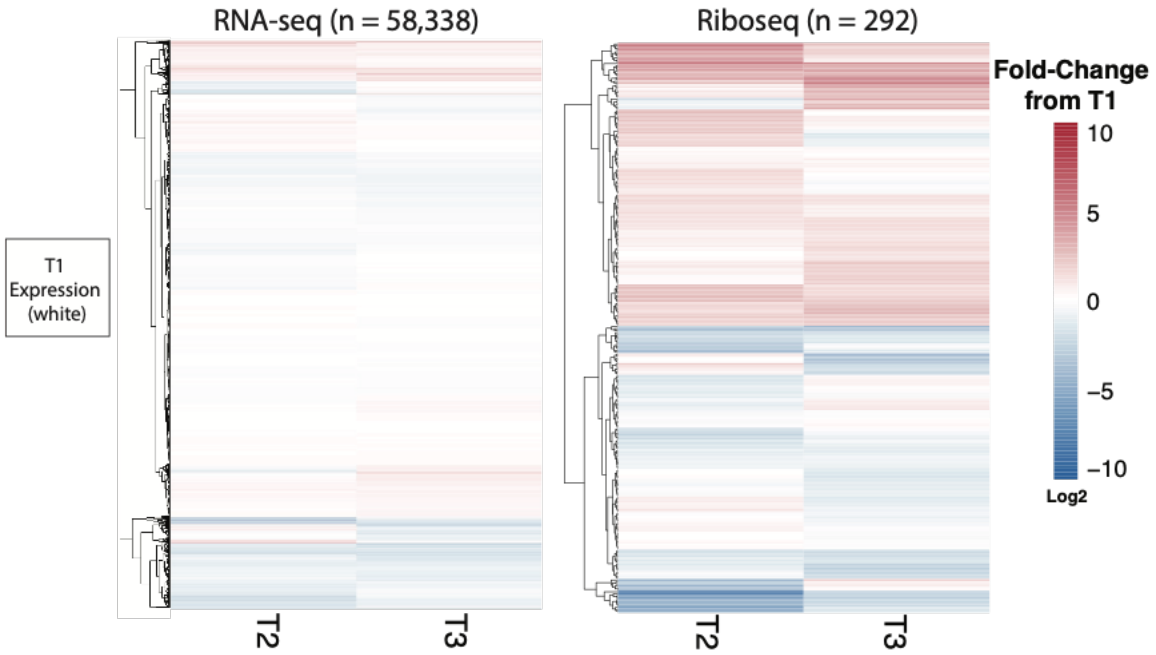


Figure xix. *Expression changes of transcripts with significant Ribo-seq mapping. Left Panel: RNA-seq expression of transcripts, Right Panel: Ribo-seq expression of transcripts*

Expression differences were compared between three key phases: Growth–Midday, Synthesis–Evening, and Division–Midnight, which together capture both the diel cycle and the cell cycle dynamics of a population that divides at night. Based on the RNA- and Ribo-seq counts, along with calculated translation efficiency, we assessed whether transcripts were 1) *buffered* (no change in translation efficiency despite a change in mRNA levels), 2) co-regulated at both the RNA and translation levels, or 3) regulated primarily at the level of translation efficiency (Figure xx, Supp. Table vi).

For all phase comparisons, there were 41,775 transcripts with significant RNA-seq mapping (> 10 reads mapped over each timepoint), but only 292 nuclear-encoded transcripts had significant Ribo-seq mapping (> 5 read over each timepoint) for comparisons (Figure xviii). Over the

comparisons, the majority of transcripts with reads mappings showed no significant change in expression in either dataset, with 53% non-significant (NS) for Growth vs Synthesis, 53% NS for Synthesis vs Division, and 57.9% NS for Growth vs Division. Transcripts displaying buffered translation were rare, comprising only 0–1.4% of the total per timepoint comparison. Transcripts showing coordinated changes in mRNA levels and ribosome occupancy represented 0-1% of the transcripts per timepoint comparison.

Transcripts regulated primarily at the translational level — showing significant changes in translation efficiency independent of mRNA levels — were the most abundant category. In the comparison between Growth and Synthesis, 17.1% and 27.7% of transcripts had decreased or increased translation, respectively. Between Synthesis and Division, 21.2% and 22.3% showed decreased or increased translation, respectively. Finally, for Growth vs Division, 23.6% and 17.1% had decreased or increased translation, respectively.

For transcripts derived from organelle genomes (the mitochondria and chloroplast), 40 transcripts were identified from the transcriptome and 13 had RNA-seq and Ribo-seq mapping for comparisons (32.5%). For Growth vs Synthesis, 7.7% were buffered with mRNA abundance up, 69.2% and 23.1% showed increased and decreased translation independent of mRNA levels. For Synthesis vs Division, 7.7% were buffered with mRNA abundance down, 23.1% and 69.2% showed increased and decreased translation independent of mRNA levels. Lastly for Growth vs Division, 7.7.% transcripts displayed coordinated changes in mRNA levels and increased ribosome occupancy, 69.2% showed NS, and 23.1% showed increased translation independent of mRNA levels.

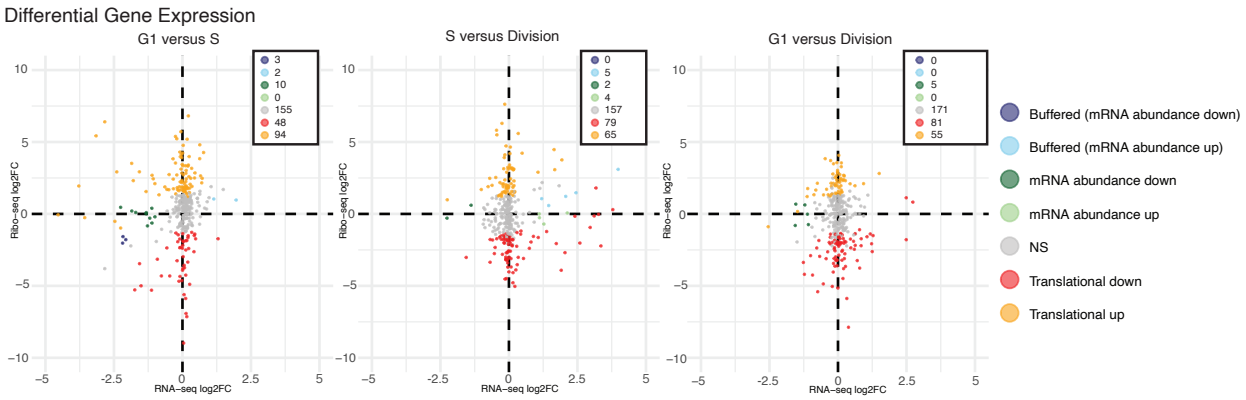


Figure xx. *Differential Gene expression on all transcripts. NS – no significant change, Buffered – mRNA counts change but the translation efficiency did not change significantly, mRNA abundance up/down -the mRNA counts and translation efficiency both changed, Translation up/down – a change in active translation without a similar change in mRNA levels.*

Analysis of differentially expressed Ribo-seq data revealed distinct functional enrichment across the three cellular states. During the Growth phase, there was a relative reduction in translation, protein folding, vesicle-mediated transport, and glycolytic and biosynthetic processes compared to the other phases, while transcripts associated with stress responses, photosynthetic electron transport, and other energy- or transport-related pathways were more highly represented (Figure xxi). In contrast, the Synthesis phase was strongly enriched for functions related to active protein production and quality control, including cytoplasmic translation, protein folding, mRNA processing, and responses to external stimuli, alongside photosynthetic electron transport and cellular responses to nutrients. The Division phase was characterized by an even stronger emphasis on ribosome biogenesis and core protein synthesis machinery, reflecting a high demand for translational capacity. Taken together, these results suggest that translation and protein folding are most active during Synthesis, ribosome production peaks during Division, and Growth emphasizes energy metabolism and adaptive response pathways while reducing direct investment in protein synthesis.

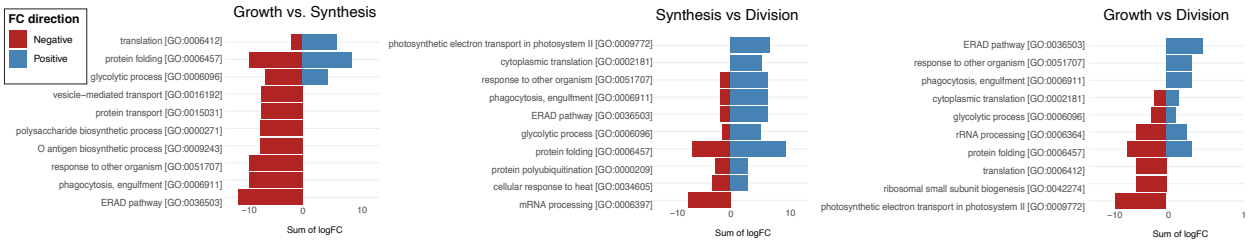


Figure xxi. *Gene Ontology (Biological Process) annotations for identifiable differentially expressed transcripts.*

Case Study: Tandem Gene Arrays and Translation

In the tandemly duplicated gene sets of *A. carterae* (described further in Chapter 5), all of the transcripts displayed differential expression at the post-transcriptional level rather than in the RNA pool (Figure xxii-A). For all of these gene sets, transcripts remained at a steady level in the RNA pool throughout the timepoints, but the ribosome occupation levels shifted over time (Figure xxii-B).

The dominant trend for these tandem gene sets was a phase-dependent increase in translation efficiency during the Division timepoint. Genes associated with chromatin packaging (major basic nuclear protein, MBNP), the mitotic and flagellar apparatus (α/β-tubulin, centrin, calmodulin), protein synthesis and proteostasis (EF1α/EF2/EF3, ribosomal protein S2, Hsp70/Hsp90), photosynthetic and respiratory energetics (e.g., psbO, chlL, chlorophyll-c complex, succinate dehydrogenase), carbon acquisition via carbonic anhydrases, and diverse transport systems ($NH4^+$, Na^+–phosphate, ZIP12, ABC transporters, permeases) exhibit increased ribosome occupation around Division relative to Growth and Synthesis. Notably, these increases frequently occur without increases in mRNA abundance, indicating regulation primarily at the level of ribosome loading rather than transcript accumulation.

These patterns support a model in which tandem gene arrays provide sustained transcript output across the photoperiod, while translational control imposes temporal specificity to coordinate

cell-cycle transitions. Elevated TE for MBNP and cytoskeletal components at Division is consistent with demands for chromatin condensation, spindle assembly, and cytokinesis, whereas TE increases in photosystem components, carbonic anhydrases, and transporters likely satisfy heightened energetic and metabolic requirements and prepare daughter cells for the subsequent light phase. Divergent TE trajectories among paralogs within some arrays further suggest emerging subfunctionalization or partitioned regulatory control at the translational level. Collectively, the data indicate that *A. carterae* leverages duplication for capacity, and translational gating for timing—coupling mitotic execution and resource provisioning to the nocturnal phase of the diel cycle.

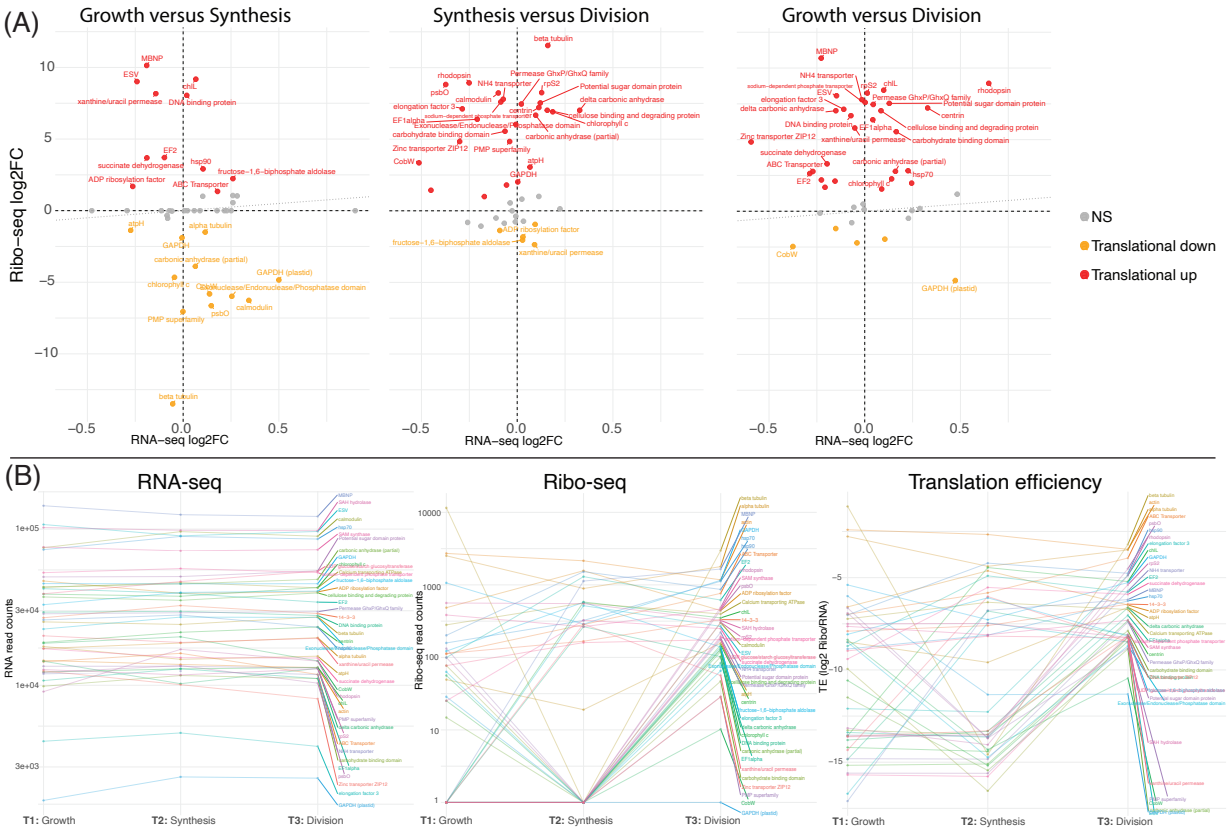


Figure xxii*. Case study: Expression for tandemly duplicated gene sets with RNA-seq and Ribo-seq. (A) Differential gene expression comparison of tandem gene sets over each timepoint. (B)*

Read counts for RNA-seq and Ribo-seq for each gene set, as well as the associated translation efficiency.

Discussion

Translation and Transcription in Dinoflagellates

Ribosome-protected (Ribo-seq) mRNA profiling in other dinoflagellate species has revealed synchronized "pulses" of translational activity across gene clusters, particularly tied to diel (light–dark) cycles, despite relatively modest variation in overall mRNA abundance (Lin 2024). This pattern—of genes being modulated at the level of ribosome engagement rather than transcription—underscores why Ribo-seq may unveil a far stronger differential response compared to RNA-seq. In this light, the observation in *A. carterae* fits neatly into a growing consensus: translational regulation, rather than transcriptional modulation, often governs gene expression dynamics in dinoflagellates.

Our Ribo-seq analysis of *A.carterae* revealed 292 differentially translated genes expressed across all 3 timepoints, whose functional enrichments across Growth, Synthesis, and Division phases reflect a dynamic, phase-specific orchestration of translational control. Specifically, the Growth phase is marked by attenuated translation of protein synthesis pathways and heightened investment in energy metabolism and stress responses; the Synthesis phase showcases peaks in translation, protein folding, and mRNA processing; and the Division phase is dominated by ribosome biogenesis and core translational machinery. This phase-structured translational program mirrors the diel translational rhythms uncovered by Morse and colleagues in other dinoflagellate systems (Bowazolo *et al.* 2022; Bowazolo & Morse 2023).

In *Lingulodinium polyedra*, ribosome profiling revealed that translation of entire metabolic pathways is temporally coordinated across three distinct times of the day: carbon fixation

enzymes peak at dawn (ZT0), DNA replication machinery at dusk (ZT12), and factors involved in transcription, translation, and amino acid biosynthesis around ZT16. These coordinated translational pulses enable metabolic specialization independent of mRNA abundance, which remains relatively stable across the diel cycle. Similarly, in the symbiotic dinoflagellate *Fugacium kawagutii*, ribosome profiling across a 24-hour light-dark cycle uncovered thousands of transcripts exhibiting rhythmic translation patterns: spliceosome and lysosome biogenesis peaked at night onset, nitrate metabolism at day onset, and ribosomal proteins toward the end of the night (Bowazolo *et al.* 2022; Bowazolo & Morse 2023).

The current observations in *A. carterae*—notably, that translation and protein-folding pathways surge during Synthesis, ribosome production is maximal during Division, and energy-related/stress response pathways dominate Growth—align convincingly with past findings. Together, these data suggest that, across diverse dinoflagellates, translational control is not merely a complementary layer of regulation but the primary mechanism by which cellular functions are temporally structured. The functional segregation we observed resonates with the rhythmically phased translational programs documented by Morse's group, reinforcing the concept that translational regulation drives metabolic and cellular phase transitions. This consistency across species underscores translational orchestration as a defining feature of dinoflagellate gene expression, adding weight to the conclusion that Ribo-seq captures biologically meaningful dynamics that are invisible to transcript-level analyses.

Diel Cycle and Cell Cycle

Three main phases of the cell cycle were aligned with the diel cycle, with Growth at midday, Synthesis appearing in the evening, and Division occurring in the night (Figure xxiii). These phases were coupled with physical measurements showing increasing activity in the cells throughout the day, with DNA measurements beginning to increase in the evening, along with

translation rates peaking by the end of the light phase, just before population doubling in the middle of the night. This trend was also followed with active translation of transcripts associated with photosynthesis and other energy pathways during the day, translation of transcripts associated with active protein production and quality control during S phase, and translation of transcripts associated with protein degradation.

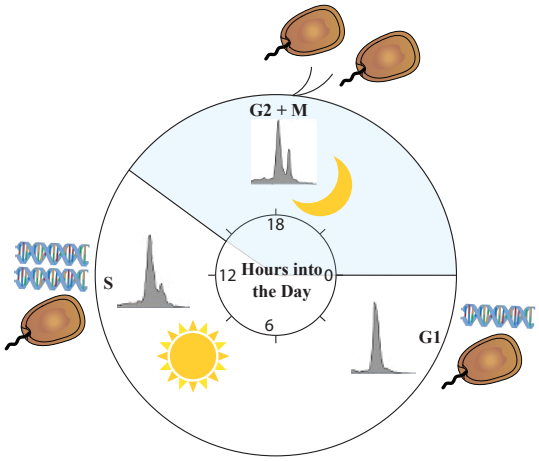


Figure xxiii. *Schematic of cell cycle phases associated with the diel cycle in A. carterae.*

The cell cycle of *Amphidinium* appears to be entrained by dark-to-light signals, with prior research theorizing that the dark/light transition cues the beginning of the cell cycle (Jia *et al.* 2019). Another paper by Gerath & Chisholm (1989) also showed a decoupling of photosynthetic capacity with the *A. carterae* cell cycle, where photosynthetic response was not associated with cell cycle phase (Gerath & Chisholm 1989). Although the photosynthetic ability of *A. carterae* cells does not seem to be dependent on cell-cycle state, our present study displayed that the stages of the cell cycle are likely fueled by energy derived during high photosynthetic times, as translation rates increased as the cells were exposed to light, with a quick decrease once the lights were turned off. Interestingly, prior research by Galleron and Durrand (1979) showed that *A. carterae* exhibits both a synchronized S-phase, as well as residual ongoing DNA synthesis

(Galleron & Durrand 1979b), or a background level of bacterial DNA synthesis, with potential bacterial species described in Chapter 2.

Dark Genes in Dinoflagellates

A considerable challenge in dinoflagellate genomics is the prevalence of "dark genes"—predicted proteins lacking recognizable identity to known sequences. In comprehensive transcriptomic analyses spanning dozens of dinoflagellate species, roughly one-third of proteins (~36%) were classified as "dark," with about 13% considered high-confidence dark proteins due to their presence in multiple taxa (Stephens *et al.* 2018). In genome-level studies, more than 50% of predicted genes in some dinoflagellates remain unannotated, with many being lineage- or species-specific (Lin 2024). Even among core, conserved gene families (e.g., in Symbiodiniaceae), dark genes often comprise over 90%, indicating their fundamental yet enigmatic nature (González-Pech *et al.* 2021).

In the current study, many differentially translated genes fall into this "dark" category—lacking robust functional annotation or sequence similarity to characterized proteins. While this complicates interpretation, it simultaneously points to novel, possibly lineage-specific functions that are translationally regulated during distinct cellular states. This underscores a broader feature of dinoflagellate genome architecture: regulation and adaptation are likely driven by uniquely evolved gene sets whose functions await elucidation. Recognizing the prominence of these dark genes emphasizes the need for integrative methods—such as structural-domain prediction, comparative genomics, and functional assays—to decode their roles.

RNA Sequencing Analyses

Integrating Illumina RNA-Seq, Nanopore Direct RNA-Seq, and Ribosome Profiling (Ribo-Seq) in a combined experimental workflow offers powerful, complementary insights across different layers of gene expression—from RNA abundance to structure and translation. These sequencing

methods are all able to measure distinct characteristics of the RNA pool within cells. RNA-seq (Illimuna) is a very high throughput sequencing method which can create a high quantity of data in the form of short and consistent reads. This method is known for being high-accuracy and cost-effective, allowing for quantification of low-expressed sequences due to the depth of coverage produced. Illumina RNA-seq is limited in that the short reads do not typically map full transcripts, leaving room for assembly error and bias. Illumina RNA-seq also requires PCR to produce the sequenced cDNA, which can introduce bias, and the system generally lacks the ability to call RNA modifications directly. The Direct RNA sequencing (ONT) directly sequences native RNA molecules, without reverse transcription or amplification, enabling sequencing of full-length transcripts with preserved RNA base modifications. This allows for more precise quantification and distinction between isoforms. Limitations of ONT Direct RNA sequencing include slightly lower accuracy compared to Illumina and less coverage than high-throughput methods involving amplification. Because of this, lowly expressed transcripts may be missed with this sequencing technology. Both of these technologies together are able to give an overview of reliable transcript presence and abundance, as well as the ability to locate RNA modifications and distinguishing isoforms.

Ribosome Profiling (Ribo-seq) is unique in that it isolated fragments protected by ribosomes, therefore not only giving a sequence, but also giving insight into the activity of the available transcripts. Unlike RNA-seq, which reflects RNA abundance, Ribo-seq directly captures translation dynamics.

Together these 3 sequencing tools provide a multi-layered, high-resolution view of gene expression from RNA abundance and characteristics to translation dynamics. In *A. carterae*, we are beginning to elucidate the ways that these layers interact to understand gene regulation. As

more is understood about the structure and functionality of the genome, specifically the numerous dark genes, more conclusions can be extracted from these datasets.

RNA Modifications

Gene expression in dinoflagellates appears to be unconventional, with heavy reliance on post-transcriptional regulation (Roy & Morse 2013; Van Dolah *et al.* 2007; Zaheri & Morse 2022), and little evidence of transcriptional differential expression (T. Chen *et al.* 2024; Lowe *et al.* 2011) . Recently, efforts have been made to understand the effect of RNA modifications on gene expression. With the current ONT, m6A, m5C and pseU modifications are able to be basecalled for m5A, m5C, and pseU, all of which are known to effect RNA expression in eukaryotes, usually associated with an increase in expression (Jiang *et al.* 2021; Nguyen-Vo *et al.* 2019; Song *et al.* 2022).

Modifications of m5C have not been explicitly characterized in dinoflagellate genomes, although orthologs of m5C writers and readers are present (Ambrosino *et al.* 2024). The m5C modification plays a role in many regulatory pathways from normal development to cellular stress responses, and in the present study this followed the same trend as the other modification where modification was correlated with increased expression (Chen *et al.* 2025).

The pseU modification is one of the most abundant in eukaryotes and plays a role in critical pathways such as gene expression regulation, translation, and splicing (Adachi *et al.* 2019). In a study by Wang and Coyne (2023), pseU synthase genes appeared to be strongly regulated in *Karlodinium veneficum*, specifically in response to algicidal treatments. Dinoflagellate transcripts are also known to carry a spliced leader (SL) RNA, a 22-nt sequence transferred from the 5′-end of a small noncoding RNA to the 5′-end of mRNA (Zhang *et al.* 2007b). In trypanosomes, SL splicing plays a key role in turning polycistronic transcripts into monocistronic ones and may also affect mRNA stability or translational efficiency (Davis 1996). The SL is also known to be

heavily modified, with a 7-methylguanosine cap followed by four additional methylated nucleotides, and a key pseU modification in a conserved position (Ullu & Tschudi 1995, 1991). These modifications are crucial for the function of the SL, and pathways regulating their modification can be used to affect gene expression. In the present study, about 5% of the transcripts exhibited high confidence pseU modification calls, with those transcripts showing significantly increased expression within the corresponding timepoint. Unfortunately, due to ONT direct RNA being read in the 3' to 5' direction, conservation of a pseU site within the SL specifically was not possible with low-coverage on the 5' ends of the reads.

In many eukaryotic systems, m6A modifications is also one of the most frequent modifications and is linked to many gene regulation mechanisms. In recent research, dinoflagellates have been found to encode the machinery to modify m6A, but measurements have shown that m1A may be more dominant. Even so, m6A and m1A modifications have been found to be much higher in dinoflagellates compared to other eukaryotes. In a study by Li *et al* 2024*, A. carterae* was found to have up to 3.05% adenines modified to m1A and only 0.63% modified to m6A (Li *et al.* 2024a). In the present study, we found that an average of 4.3% of adenines having the m6A modification. Calls for m6A modifications were also correlated with higher expression, but patterns of differential expression were not consistent across timepoints. Although ONT modification calling is becoming better, there is a possibility that m6A and m1A modifications are both being counted, as the dorado basecaller is only trained on m6A, not m1A. Insight into the distinction between these modification types is still being researched.

Polycistronic transcription

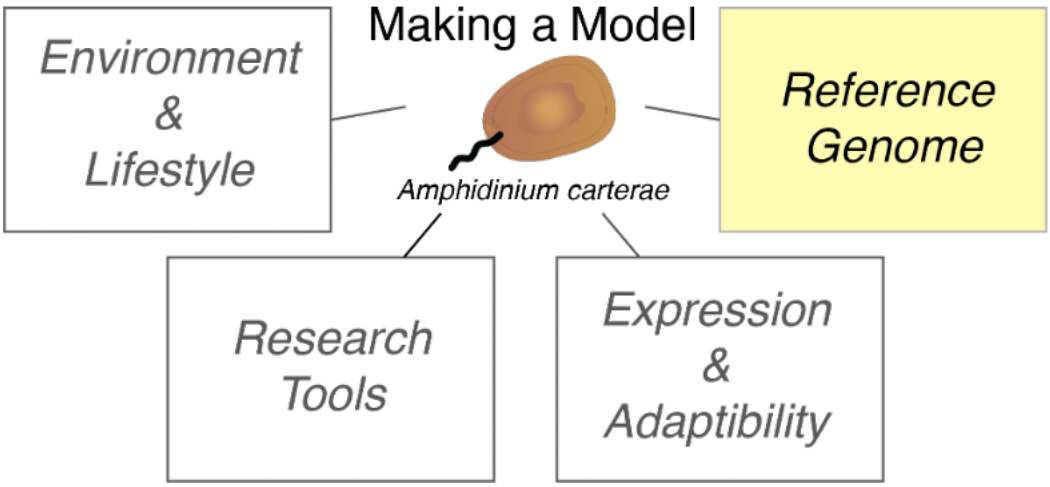
The theory of dinoflagellates using polycistronic transcription was prompted by the discovery of both tandemly repeated gene arrays and the SL sequence on RNAs. This is similar to gene structures found in Trypanosomes, where RNA polymerase II transcribes long stretches of genes

as polycistronic transcription units (PTUs) —multiple adjacent genes are transcribed together in a single long RNA precursor (Clayton 2019). These long precursors are processed into individual, mature mRNAs through spliced leader (SL) trans-splicing and polyadenylation at each gene's end. This arrangement minimizes regulation at the level of individual gene transcription. Instead, using post-transcriptional regulation, controlling gene expression via mRNA processing, stability, translation, and decay. Because dinoflagellates also display heavy reliance of post-transcriptional regulation, the theory of polycistronic transcription appears to fit in this perspective. But recent studies have shown a lack of evidence of this mechanism occurring – for example, Beauchemin *et al.* (2012) directly tested whether tandem gene arrays in dinoflagellates produce a single, polycistronic transcript. The results showed that these transcripts are highly conserved and not polycistronic in nature (Beauchemin *et al.* 2012). In the present study, the long-read direct RNA sequences did not traverse intergenic spaces between gene copies, showing that polycistronic transcription is likely not occurring in these regions.

Conclusions

Dinoflagellate gene regulation is still an evolving field with many potential mechanisms being part of the regulatory system, from base modifications to gene structure. In our study, we present multiple new methods for observing changes in the transcript pool. New ONT basecalling and modification detection now allows for not only sequencing full-length transcripts but also understanding the distribution of their modifications. Although models are still being developed, this emerging field is rapidly progressing to allow for more distinct predictions. We also found that ribosome profiling may be a better way to understand shifts in expression, as RNA pools appear more constant than actively translated transcripts. Together, these methods begin to give insight into the effects certain mechanisms have on gene expression in the enigmatic dinoflagellate.

Chapter 5: Collectors, not hoarders: Complex gene structures in *Amphidinium carterae* revealed through Nanopore sequencing[3]



Dinoflagellate Genomics

Dinoflagellate genomes are estimated to be very large, with some nuclei measurements alluding to DNA content over hundreds of picograms per cell, compared to the much lower DNA content of 6.5 pg in *Homo sapiens*, 0.44 pg in *Arabidopsis thaliana,* or 0.12 pg in *Chlamydomonas reinhardtii* (Chaux-Jukic *et al.* 2021; Lin 2011; Schmuths 2004; Sun *et al.* 2018). The dinoflagellate genome contains a high abundance of the otherwise rare, modified base, 5-hydroxymethyluracil, which may affect chromatin accessibility, mobile elements, and gene expression (Li *et al.* 2024b; Marinov *et al.* 2021). The dinoflagellate chromosomes also remain condensed during interphase in a refringent "liquid crystalline" state and lack conventional nucleosomes (Spector 1984). Many dinoflagellate genes are organized in multiple repeats as tandem arrays, which can be intron-rich with short exonic regions (Bachvaroff & Place 2008; Beauchemin *et al.* 2012). Horizontal gene transfer is also a common occurrence for dinoflagellate genomes, adding to their large and complex genomes (Wisecaver *et al.* 2013).

[3] Results of this chapter have been accepted by *BMC Genomics* (2025) under the same title.

The expansion of dinoflagellate genomes has been theorized to be due to various mechanisms including duplication events, retro-transposition, and rampant horizontal gene transfer (Lin 2024). Because of dinoflagellates' observed affinity for gene acquisition, they are colloquially known as gene “hoarders” - generally accepting genes into their nuclear genome at a relatively high rate (Dorrell & Howe 2012; Moszczyński *et al.* 2012; Williams *et al.* 2021; Wisecaver *et al.* 2013)**.** This is particularly emphasized in research regarding tandem duplications, where some gene copies have been observed in the thousands (González-Pech *et al.* 2021; Le *et al.* 1997). The presence of this extreme duplication raises questions related to the evolutionary advantage for retention of these expanded gene copies. Here, we present an alternative to the gene “hoarder” theory, which assumes unfiltered gene retention; instead, we present a theory of gene “collection” in which dinoflagellates use an alternative mechanism of expanding gene copy number to affect gene expression in a system where transcriptional regulation is atypical (Roy *et al.* 2018; Roy & Morse 2013)**.** This gene “collector” theory was established on the basis that gene duplication events occur due to evolutionary pressures and life history strategies and directly corresponds to gene expression (Lazo-Murphy & Peng 2025).

The control of gene expression is ambiguous, as dinoflagellates seem to rely heavily on post-transcriptional regulation (Van Dolah *et al.* 2007), while there is also evidence of some select genes under differential expression (Babaei *et al.* 2024; T. Chen *et al.* 2024; Lowe *et al.* 2011). Prior research has also shown that transposable elements (TEs) may play a large role in gene expression, from contributing as regulatory elements to acting as non-coding RNAs that modulate gene expression post-transcriptionally (Drongitis *et al.* 2019)**.** This may work in conjunction with the use of modified bases, such as 5-hydroxymethyluridine, operating as epigenetic markers for TEs (Li *et al.* 2024b). Other theories as to how genes are expressed include the use of a dino-

specific, 22-nucleotide *trans*-spliced leader, which is donated to the mRNA and may play some role in regulation (Jones *et al.* 2024; Roy *et al.* 2018). The use of alternative eukaryotic translation initiation factors has also been suggested as a potential mechanism for regulating the initiation of translation, which may be necessary in a system where transcription remains relatively constitutive (Jones *et al.* 2015).

Dinoflagellate genomic research is at an early stage, with completely assembled reference genomes still limited, primarily to symbiotic and free-living species from the Symbiodiniaceae family, which possess reduced genome sizes relative to other core dinoflagellates (González-Pech *et al.* 2021). Advances have recently been made in sequencing *Polarella glacialis*, a free-living relative of Symbiodiniaceae, utilizing PacBio long-read and Illumina short-read hybrid approaches (Stephens *et al.* 2020). Emerging *de novo* assemblies suggest that core dinoflagellate genomes may differ from previous flow-cytometry genome size estimations based on fluorescence, with current assemblies ranging from <0.1 Gb in parasitic syndineans to over 250 Gb in free-living core dinoflagellate species (Lin 2024; Shah *et al.* 2024). One strain of *Amphidinium carterae*, a basal species of core dinoflagellate, has recently had its genome assembled at approximately 1.2 Gb and scaffolded into 34 pseudo-chromosomes (Li *et al.* 2024b). Here we build upon the genomic resources for *Amphidinium* with the present assembly and annotation to assess intra-species retention of genomic features.

Dinoflagellate genome annotation is challenging, in part due to unique genomic architecture, such as vast repetitive regions, non-canonical intron splicing, and unique stop-codon usage. Deviations from traditional Eukaryotic gene structures may cause common gene prediction programs to fail (Aranda *et al.* 2016; Chen *et al.* 2020; Y. Chen *et al.* 2024; Farhat *et al.* 2021)**.** Furthermore, a significant proportion of predicted proteins have no recognizable function or orthologs, termed "dark genes" (Stephens *et al.* 2018)**.** This high sequence divergence undermines homology-based

annotation tools. Protein prediction tools can also lead to biases based on the prediction tool used, which can have an impact on subsequent comparison and analysis (Chen *et al.* 2020)**.** These prediction tools may commonly miss tandemly repeated genes or classify them as non-coding repeats. Lastly, RNA-editing is widespread throughout dinoflagellate transcripts, with between 7.5-13.2% of dinoflagellate transcripts undergoing RNA-editing, which impacts coding sequences and confounds gene predictions based on genomic DNA (Y. Chen *et al.* 2024)**.** Thus, the choice in annotation method is crucial, here we used a well-curated transcriptome-guided method for annotating the present assembly**.**

In the present study, we generated the first deep-sequenced genome assembly of *A. carterae* using only Oxford Nanopore Technology (ONT) long-read sequencing. This species was selected for its basal phylogenetic position, broad environmental distribution, and the production of bioactive amphidinol toxins, which is relevant to dinoflagellate research as toxin-production is one of the main drivers of ecological and economic disasters (Haq *et al.* 2017; Zhang *et al.* 2007a). Crucially, we present a comprehensive transcriptome-guided annotation of tandemly repeated genes for *A. carterae*, as well as gene models, tRNA identification, and functional annotations. Assembly completeness was evaluated using BUSCO scores, super-transcriptome mapping, and read coverage analysis. This annotated genome represents the novel use of ONT sequencing to assemble highly complicated genomes and offers a significant resource for investigating dinoflagellate genome structure and gene content to provide a foundation for future functional and evolutionary studies.

Methods

Cell Culturing

Amphidinium carterae NCMA 1314, provided by the National Center for Marine algae and Microbiota, was grown in a defined, enriched artificial seawater (ESAW) marine media with a

salinity of 20 supplemented with f/2 nutrients (without silicates). Cultures were made at a starting concentration of 10,000 cells mL^{-1} and allowed to acclimate in a 14:10 h light–dark period for 14 days in 100 µmol photons m^{-2} s^{-1} at 25°C (Berges *et al.* 2001). The cultures were continuously grown with 100 µg/mL carbenicillin, 50 µg/mL kanamycin A, and 50 µg mL^{-1} spectinomycin (Liu *et al.* 2017).

Nanopore long-read sequencing & assembly

The same sequence data as described in a previous publication on the microbiome was used to assemble the genome of *A. carterae* (Judd *et al.* 2024). Briefly the DNA was extracted and sequenced using a combination of Nanopore PromethION and MinION platforms, and then basecalled using the "super" accurate model with dorado (7.2.13) using the dna_r10.4.1_e8.2_400bps_sup@v4.3.0 model. After removing sequences that mapped to the bacterial reference genomes and the ONT DNA control strand, the remaining genuine *A. carterae* sequences were pooled from multiple samples and assembled with the Flye *de novo* assembler (2.9.1–b1780) (Kolmogorov *et al.* 2019). Bacterial sequences, described by Judd *et al* 2024, were excluded from the dinoflagellate genomic analysis (Judd *et al.* 2024). After assembly, contigs were further filtered by length (10 kB) and coverage (30X). Lastly, the absence of duplicated contigs was confirmed by running a deduplication script from the MaSuRCA genome assembly package (Zimin *et al.* 2013).

Gene Annotation

Gene annotation was performed using several complimentary methods to ensure thorough, valid identification (Figure xxiv). First, transcripts created with RNA-seq data collected by Bachvaroff *et al* (2014) were assembled using Trinity and compressed into super-transcripts, which were filtered based on TPM using Bowtie2 alignment and quantified with StringTie (Bachvaroff *et al.* 2014; Langmead & Salzberg 2012; Pertea *et al.* 2015)**.** BLASTn was used to identify tandemly

repeated genes within the genome with a minimum percent identity of 95% and query coverage of 80% (truncated copies were identified as 'pseudogenes'). Tandemly repeated genes were identified as super-transcripts which mapped at least 3 times in a row to one contig, with consistent directionality. After BLASTn identification these regions were then manually verified and annotated with Apollo using RNA-seq read alignments, exonerate results, and CDS sequence identity to curate stop/start codons and intron-exon boundaries to create ORFs (Dunn *et al.* 2019). Tandem genes were masked prior to the repeat masking step performed with Dfam TE Tools's RepeatMasker (Tarailo-Graovac & Chen 2009). Complimentary with the manual curation of tandem genes, the super-transcripts were mapped to the genome using MAKER and exonerate, with intrinsic flexibility allowing for the canonical GC-AG and GA-AG dinoflagellate splice sites (Cantarel *et al.* 2008; Davidson *et al.* 2017). Exonerate gene predictions and the manually verified tandem gene sets were merged and deduplicated into one comprehensive annotation**.** Predicted proteins with repeated domains were excluded from the tandem gene group because of domain shuffling, which represents a different type of duplication within domains that were often not exact copies. These included the anomalous RuBisCo and LHC genes, which were both previously described as polypeptides containing multiple mature proteins derived from a single mRNA (Hiller *et al.* 1993; Rowan *et al.* 1996).

Mitochondrial contigs were identified by BLASTn hits to mitochondrial genes and verified by their high AT content (Jackson *et al.* 2007; Lin *et al.* 2002)**.** To identify tRNAs, trnascanSE and infernal were used with default parameters, and BLASTn with a percent identity minimum of 95% was used to identify 18S, 23S, 5.8S, 5S and the DinoSL genes (Chan *et al.* 2021; Nawrocki & Eddy 2013). BLASTn with the 22-base SL and exact matches was used to identify DinoSL genes, using the well-conserved dinoflagellate SL sequence "ACCGTAGCCATTTTGGCTCAAG-GTACACATTGGG" with the hyphen indicating the splice site (Jones *et al.* 2024; Zhang *et al.* 2007b)**.**

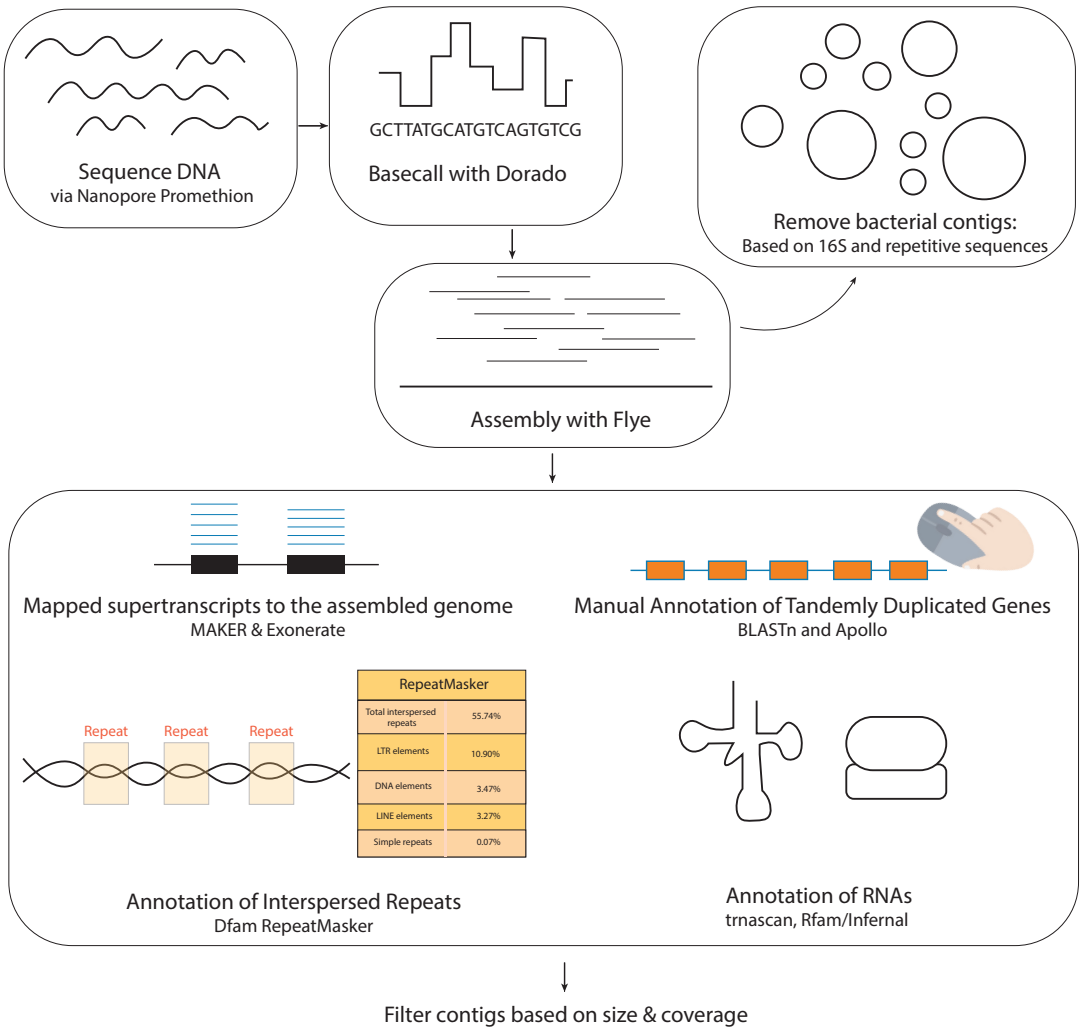


Figure xxiv. Genome assembly and annotation pipeline

Assessing Annotation Completeness

To assess completeness of the genome, the *A. carterae* transcriptome was compared to the genome assembly, with BLASTn using a percent identity cut-off of 97%. Transcripts with >50% of their length mapping across one contig were then quantified. This was also performed with the compressed Trinity-assembled super-transcriptome. Mapping of the regular transcriptome and the super-transcriptome was done to highlight the limited differences between these datasets.

Measuring Selection

For each set of tandemly repeated genes the nucleotide and amino acid differences between copies was measured using a custom python script incorporating Biopython's Distance Calculator based on the aligned CDS and peptide sequences for each gene set (Cock *et al.* 2009). For the MBNP gene family, complete alignments of each gene (including the intron) from the splice acceptor to poly-A site were constructed and the exact nucleotide differences were quantified. A phylogenetic comparison of the MBNP gene copies was done using aligned nucleotide sequences using RaxML (Kozlov *et al.* 2019). All nucleotide and amino acid logos were created by importing aligned nucleotide sequences through the R "ggseqlogo" package (Wagih 2017).

To further investigate the selective pressures acting on all of the tandemly repeated genes, the nucleotide and amino acid sequences were aligned, and codon alignments were subsequently created using a perl script from the Phylogenetic and Molecular Evolution (PhaME) analysis tool (Shakya *et al.* 2020)**.** Newick files were then created with FastTree and site-specific codon analysis was performed using the Fixed Effects Likelihood (FEL) method. This analysis estimated the synonymous (α) and non-synonymous (β) substitution rates for each codon position across all gene copies.

Gene Expression

Gene expression was based on RNA-seq read mapping using Bowtie2 with secondary mapping disabled to the super-transcripts, which were in turn matched to the assembly using exonerate. Samtools coverage calculations and FPKM values from Stringtie using a super-transcript-based estimate agreed closely with mapping reads to the genome with hisat2 and Stringtie abundance estimate methods.

Telomere Identification

Telomeric regions of the contigs were identified using BLASTn of the 7-base repeat motif "TTTAGGG"**,** with a word-size of 7 and DUST filtering and soft-masking disabled (Kwok *et al.* 2024). The identified regions were further filtered to start at most 100 bases from the end of the contig and have a repeat length minimum of 1 kb.

Comparing References

The present assembly was compared to the 2024 Li *et al* assembly of another strain of *A. carterae* using the minimap2 tool in the D-genies suite, as well as QUAST (Cabanettes & Klopp 2018; Gurevich *et al.* 2013; Li *et al.* 2024b). The present, curated set of tandem gene annotations were further identified in the Li *et al* 2024 assembly using BLASTn with a cutoff of 95% and minimum query coverage of 70% after parameter optimization.

Conserved tandem genes actin and hsp90 were also identified across references of other dinoflagellate species by comparing the *A. carterae* actin transcript to Trinity assembled transcriptomes of both Symbiodiniaceae and *Polarella*, identifying the most likely actin-candidate within these transcriptomes, and blasting the aligned actin regions to the other dinoflagellate assemblies with percent identity >90% and with query coverage cutoffs set by species-actin alignment percents (González-Pech *et al.* 2021; Stephens *et al.* 2020).

Results

Assembly

After bacterial reads were filtered from the data, this left 6.7 M putative *Amphidinium* reads (88 Gb, 21 kb N50, Supp. Table vii) which were assembled into a purified dinoflagellate genome totaling 1.25 Gb across 2,514 contigs (Figure xxv, Table v). Of these, 36 are mitochondrial (See black bubbles in Figure xxv-A), and have a total length of 642 kb. These mitochondrial contigs

have an average contig length of 17 kB and a AT percentage of 70%. The total coverage, or depth of reads per base, for the mitochondrial contigs ranges widely from over 10,000 reads per base to under 100. The remaining 2,478 contigs have an average length of 502 kb, an average GC content of 44% (range 40-50% GC), as well as an average coverage of over 60X with a few outliers having very high coverage. The assembly had a maximum contig length of 7.7 Mb, an N50 of 1.1 Mb, and an L50 of 329 bases. The contigs were categorized into 3 groups based on their annotated features: 1,743 have mRNA (872 of these have no other RNA elements) (1.2 Gb), 111 have only other RNA elements (not mRNA) (11.8Mb), and 624 are non-coding contigs with no annotated RNA elements (30.4Mb) (Figure xxv). Overall, 50,641 mRNA features from the Trinity-assembled super-transcriptome are annotated, most based on exonerate comparisons to the super-transcriptome, with the remaining genes being tandemly repeated gene sets.

Table v. Genome characteristics

Genome	Total Length	No. of contigs	GC %	Average Contig Length	Median Contig Length	Average Coverage	Median Coverage	N50	L50
Nuclear	1,253,139,580	2,478	44.18	505,706.05	237,667	60.44	55	1,146,910	329
Mitochondrial	642,647	36	30	17,278	17,326	2,404	592	19,139	15

Non-coding RNA elements

The 111 contigs that have only non-mRNA annotations include tRNAs, ribosomal RNAs, and microRNAs (11.8Mb). In total, there are 7,462 tRNAs as well as 1,593 pseudo-tRNAs. This includes 23 unique tRNAs including modified versions such as i-Methionine. For the ribosomal RNA, there are 8 SSU and 7 LSU mostly clustered on one contig. Splice-leader (SL) genes have been relatively well-documented in dinoflagellate genomes as small non-coding RNAs that transplant a small RNA fragment to the 5’ end of premature RNA molecules (Lidie & Van Dolah

2007). In the current assembly, the SL-gene appears 180 times. There are also 441 5S genes, 54 of which appear on the same contig as the SL genes. The U2, 3, 4 and 6 RNA are present with 8, 20, 3, and 78 copies, respectively. A single site of 24 RNAse Mitochondrial RNA Processing (RF00030) repeats is also found on one contig.

Transcriptome Mapping to the Genome

The assembled contigs contain most of the known transcripts in *A. carterae* based on the comprehensive Trinity-assembled transcriptome mapping. Most (96%) of the transcripts were found to map back to the assembly. Once the transcript isoforms were compressed into 56,775 super transcripts, 96.5% of the super transcripts were found within the assembly. Based on these results, there appears to be a near complete mapping of the previously sequenced *A. carterae* transcriptome to the assembled genome.

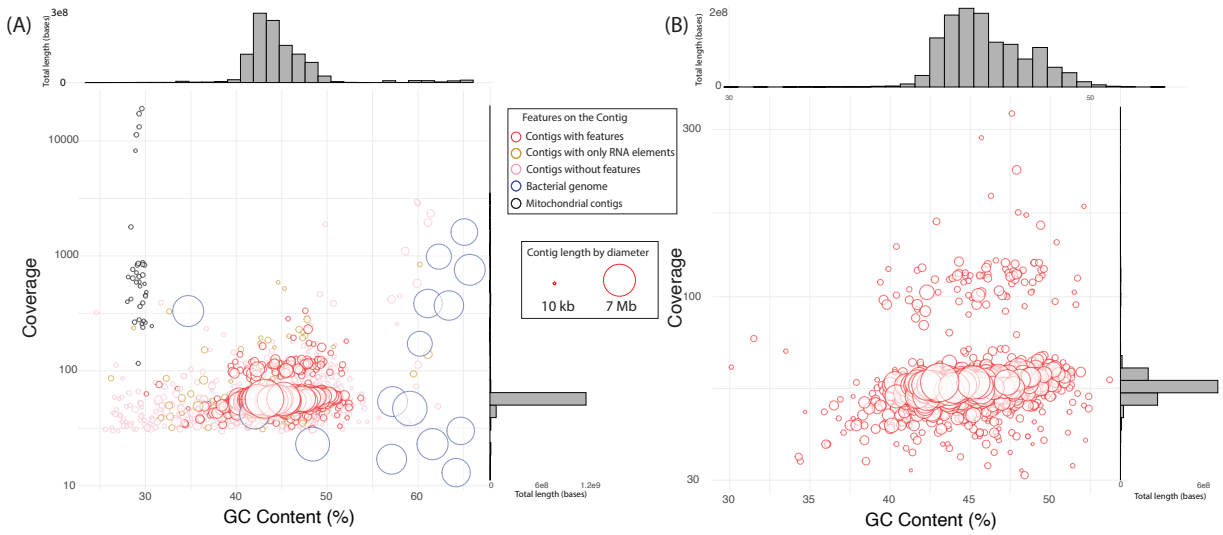


Figure xxv. *Genome Contig characteristics. (A) All contigs are plotted and divided into mitochondrial, bacterial, non-coding, with RNA elements only, and coding (with mRNA). (B) Bubbles include only contigs with coding features, with mitochondrial and bacterial contigs removed. Bubble size corresponds to contig length.*

Repetitive Regions

Repeats in the genome cover 55% of the assembly or approximately 720 Mb (Figure xxvi-A). These repeats include 3.27% LINEs, 10.90% LTR elements, 3.47% DNA elements, 1.94% simple repeats, 0.07% low complexity repeats, and 37.10% of repeat families that were not previously identified. The unclassified repeats contain 1,964 repeat families, with the largest family covering 21.8 Mb bases of the assembly. Repeats cover on average 62% of the nuclear contigs, with repeats covering about 90% of contigs without identified coding regions and covering about 56% for contigs with identified coding regions (Figure xxvi-B).

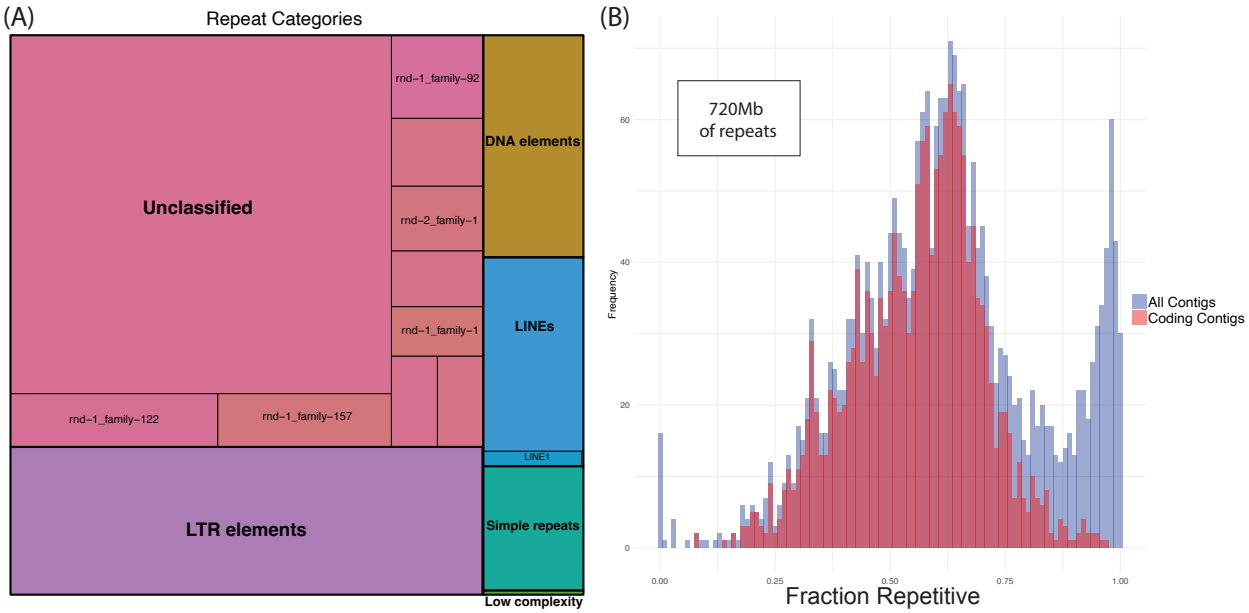


Figure xxvi. *The 720 Mb of repetitive sequence found within the assembled A. carterae genome. (A) Classification of repeats found in the genome with area proportional to the number of bases. (B) Percent of contig length occupied by the identified repeat regions.*

Gene Structure

In the nuclear genome, 50,641 transcripts were annotated in the assembled contigs, and 91.2% appear only once. There are 384,844 total exons found within the mapped transcript genes (an average of 7.6 exons per gene). Exon length ranges from 3 to 11,218 bases with a median of 74

(Figure xxvii-A). Introns are generally longer than exons, with a median length of 439 and a maximum of 87,555 bases, and their frequency varies substantially from gene-to-gene (Figure xxvii-B). Only 19,702 (38.9%) mapped transcripts appeared as single exon genes in the genome, with the largest number of exons-per-gene being 133 (Figure xxvii-C). Splice sites show mostly canonical GT with some non-canonical splice sites using GC/GA donors and nearly universal AG splice acceptors as well, a pattern previously described in dinoflagellates (Figure xxvii-D) (Beedessee *et al.* 2020; Rowan *et al.* 1996).

Nested Genes

Many contigs exhibit gene arrangements spanning large genomic regions, characterized by short exons and long introns, with additional genes embedded within these introns (Figure xxvii-E) compared with tandemly duplicated genes (Figure xxvii-F). Genes inside of introns of other genes has been called ‘nested’ where genes are found nested in introns within a longer ‘host’ gene (Kumar 2009). Nested genes are common in the assembly, with 26.1% of identified genes nested within a host gene intron, of which about 10% are nested even further. The host gene with the largest number of nested genes is on contig 4281, which has 28 nested genes across 47 kB of a contig. The average FPKM of nested genes is 2.07 (SD: 5.7) times higher than that of their corresponding host genes, with a large distribution of nested:host expression ratios. When filtered for high expression (FPKM > 10) for both host and nested genes, the nested gene’s expression ranged from 0.07 to 14 times that of the associated host genes. Also, 40.2% and 29.1% of nested genes had higher or lower expression compared to the host genes, respectively, with 30.8% having similar expression (within 75%). Overall, the presence of these nested genes displayed how different regions of the contigs experience differential transcription.

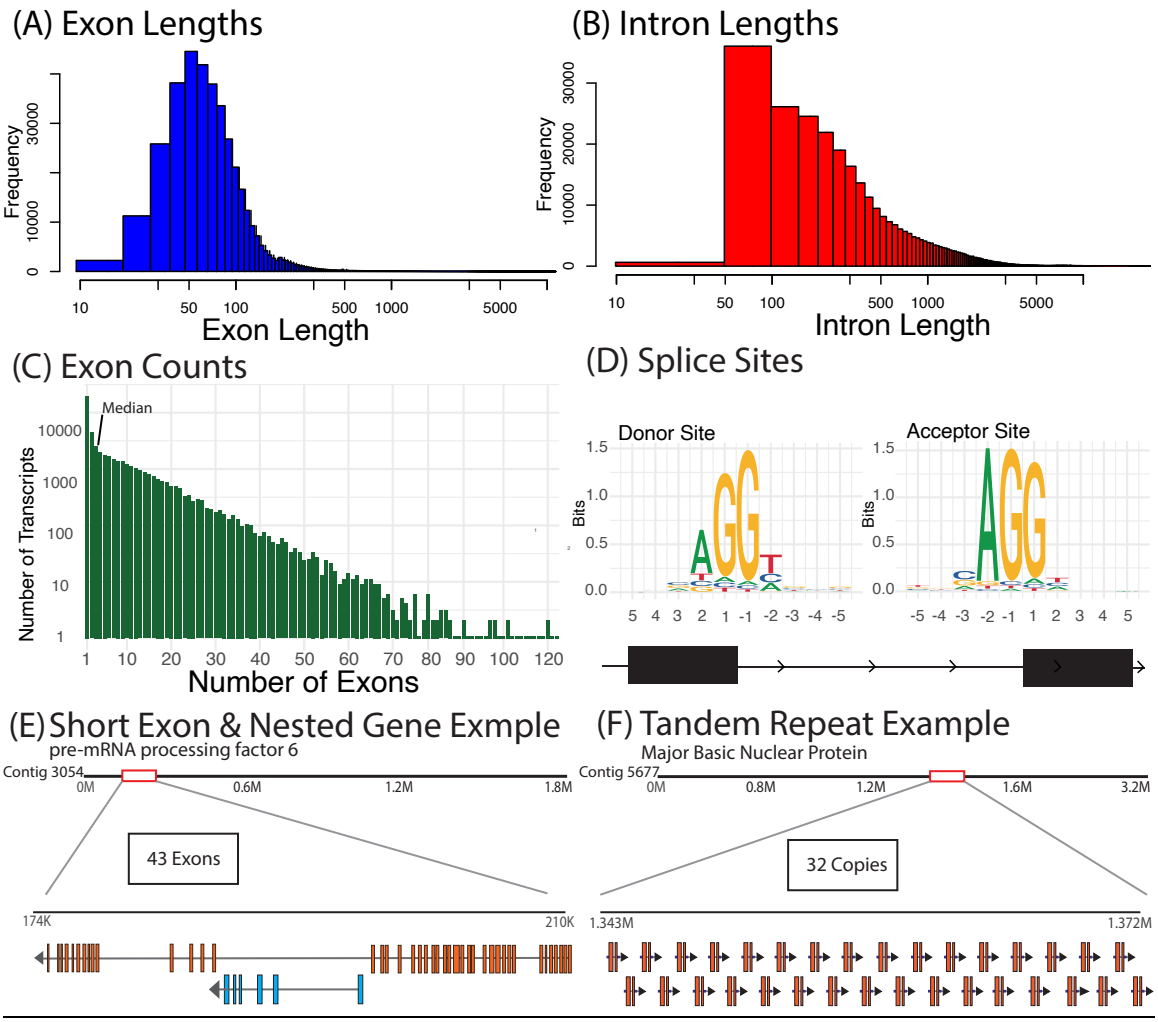


Figure xxvii. *Gene features in A. carterae: (A) Histogram of exon lengths, (B) Histogram of intron lengths. Histogram binning was determined by the Freedman-Diaconis rule. (C) Bar graph of the number of exons per transcript mapped on the genome, (D) Consensus logo of splice sites with the size of the nucleotide being the proportion of exons with that base occurring at the splice location, (E) An example of a common gene structure in A. carterae with many short exonic regions interrupted by long introns in the pre-mRNA processing factor 6, which is an Alveolata BUSCO gene (Group 4205at33630) recovered by transcript mapping. This gene spans ~36 kb of a 1.9 Mb contig. There is a six-exon gene nested within the larger intron. (F) An example of a tandemly repeated gene: Major Basic Nuclear Protein (MBNP) forms a large*

cluster of 32 repeated genes which span ~40 kb of the 3.2 Mb contig and have a single short exon at the same site for each copy, a longer second exon, and a compact intergenic region between copies.

Genome BUSCO

BUSCO analysis of the assembled *A. carterae* genome using AUGUSTUS in genome mode initially reported only 58.5% completeness (53.2% single-copy, 5.3% duplicated, 8.2% fragmented, 33.3% missing). In contrast, analysis of the super-transcriptome indicated much higher completeness (93.6% complete, 3.5% fragmented, 2.9% missing), and 52 of the 57 BUSCO genes absent from the genome were recovered in the transcriptome. Two of the remaining BUSCO genes lacked any dinoflagellate reference sequences. Of the 52 transcriptome-recovered genes, 39 mapped to the genome by BLASTn, indicating that they were present but not annotated by the BUSCO-AUGUSTUS pipeline. Likewise, the 14 genes flagged as fragmented in the genome were fully recovered as transcripts. Incorporating transcriptome-supported annotations raised the genome completeness estimate to 89.6% (82.5% single-copy, 5.3% duplicated, 1.8% fragmented, 10.4% missing). Of the genes still missing, 12 (67% of the missing set; 7% of the total BUSCO set) were nevertheless present in the transcriptome (Supp. Table viii).

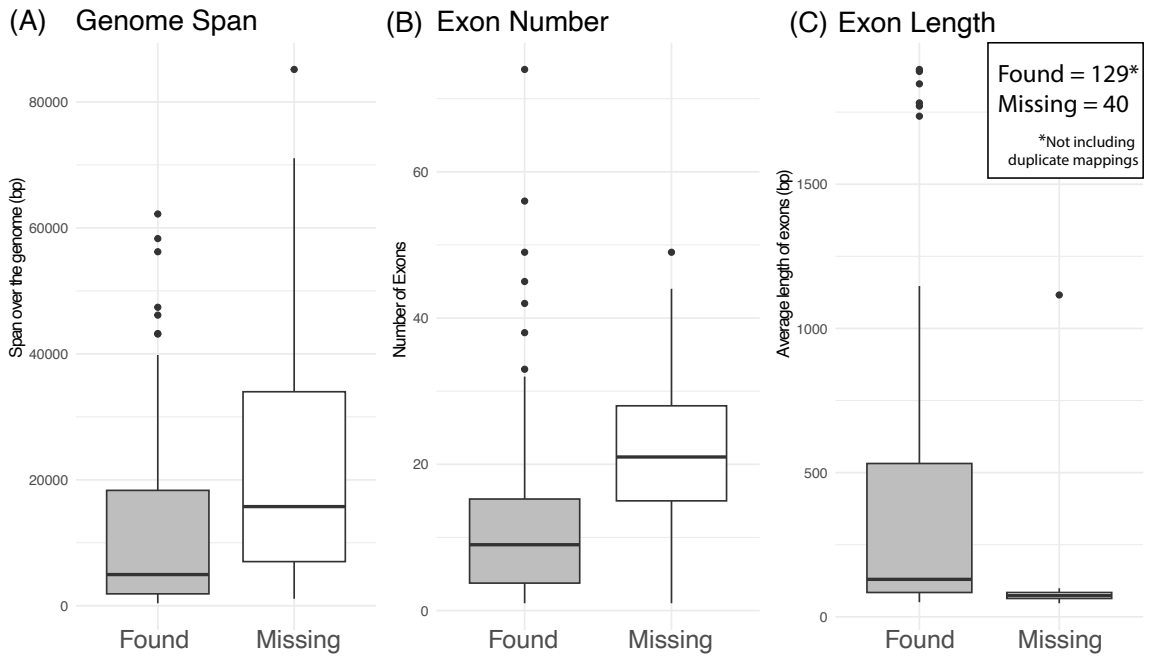


Figure xxviii. *Characteristics of the genes found using AUGUSTUS compared to the genes initially 'missing' (before transcriptome matching with BLASTn) in the assembled A. carterae genome. (A) Genomic span of the identified BUSCO gene, (B) number of exons per gene, and (C) the average exon length per gene.*

The BUSCO genes missed by AUGUSTUS typically span a longer region of the genome with an average size in the genome of 24 kb and a maximum of 85 kb when including introns, while those found with the standard BUSCO approach had an average size in the genome of 12kb and a maximum of 62 kb (Figure xxviii-A). The BUSCO genes that were missing from the genome analysis, but present as mapped super-transcripts, had an average of 22 exons with an average length of 100 bases, while the genes initially found with the standard BUSCO approach had 12 exons on average with an average length of 388 bases (Figure xxviii-B and xxviii-C).

Tandemly repeated genes

Tandemly repeated genes are a common feature of the *A. carterae* genome (Figure xxvii-F, Supp. Table ix). These sets are defined as genes repeated a minimum of 3 times on a contig with

consistent directionality. Overall, 68 different genes are in tandem arrays, with copy numbers ranging from 3 to 52 (Figure xxix-A). Many, but not all of these genes also contain high identity ‘singletons’ that are found as single genes on different contigs from the arrays. The tandem genes have an average of 2.57 exons with average exon and intron lengths of 1,043 and 514 bases, respectively (Supp. Figure vi).

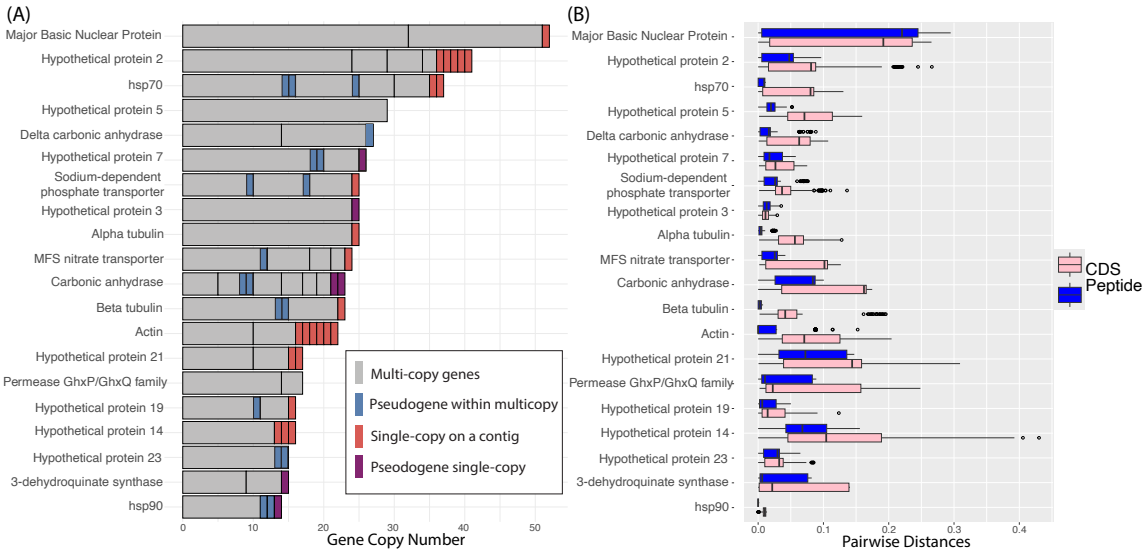


Figure xxix. The top 20 most duplicated tandemly arranged genes in the A. carterae genome. *(A) Copy numbers of the tandemly repeated genes identified in the assembled A. carterae genome with each division within the bar representing a different contig. (B) Pairwise distances within tandem gene sets comparing nucleotide differences with amino acid differences between copies. Truncated genes were removed from this analysis.*

The amino acid variation within these tandem repeats is minimal, with the average pairwise protein sequence distance being 0.066 (Figure xxix-B). The CDS variation between these repeats is slightly greater, with the average pairwise distance of 0.076 and a broader range of values. The larger distances between the MBNP repeats are due to a unique variable gene structure for this category detailed below.

A Case Study: Major Basic Nuclear Protein

The Major Basic Nuclear Protein (MBNP) also referred to as "Histone-Like Protein" or HLP is a good example of tandemly repeated genes because it is the most highly duplicated gene within the tandem gene families, containing 52 copies with high overall expression (FPKM of 6,044). In core dinoflagellates this protein is the most abundant protein complexed with DNA and is a short 108-111 amino-acid (327-336 CDS) long protein with many basic amino acid residues (Chudnovsky *et al.* 2002; Sala-Rovira *et al.* 1991). Previous studies showed that *A. carterae* has two tandemly duplicated variants of MBNP (Janouskovec *et al.* 2017). Here we have identified and characterized the two tandemly repeated types where the longer variant has a single 3 AA insertion in the first exon near the start codon when compared with the short form, and the two variants are generally present as long and short pairs in contig 5677 or short and long pairs in contig 4958, which are defined as two tandemly repeated groupings (Figure xxvii-F) leading to larger distance estimates as shown in Figure xxx-B. Finally, there is a lone long copy on contig 631 which shares the same intron-exon structure as the other genes. Within both MBNP forms, the 5' UTR, first exon, second exon, and 3' UTR remain relatively well conserved with a maximum of 34, 66, 18, and 62 nucleotide differences found in direct copy comparisons, respectively (Figure xxx-A). This variation increases in the intronic and intergenic regions to 103 and 111 maximum nucleotide differences, respectively. Nucleotide differences are even less for all parts of the gene (5'UTR, exons, intron, and 3'UTR) when calculated within variant groups, although the intergenic regions remain more variable.

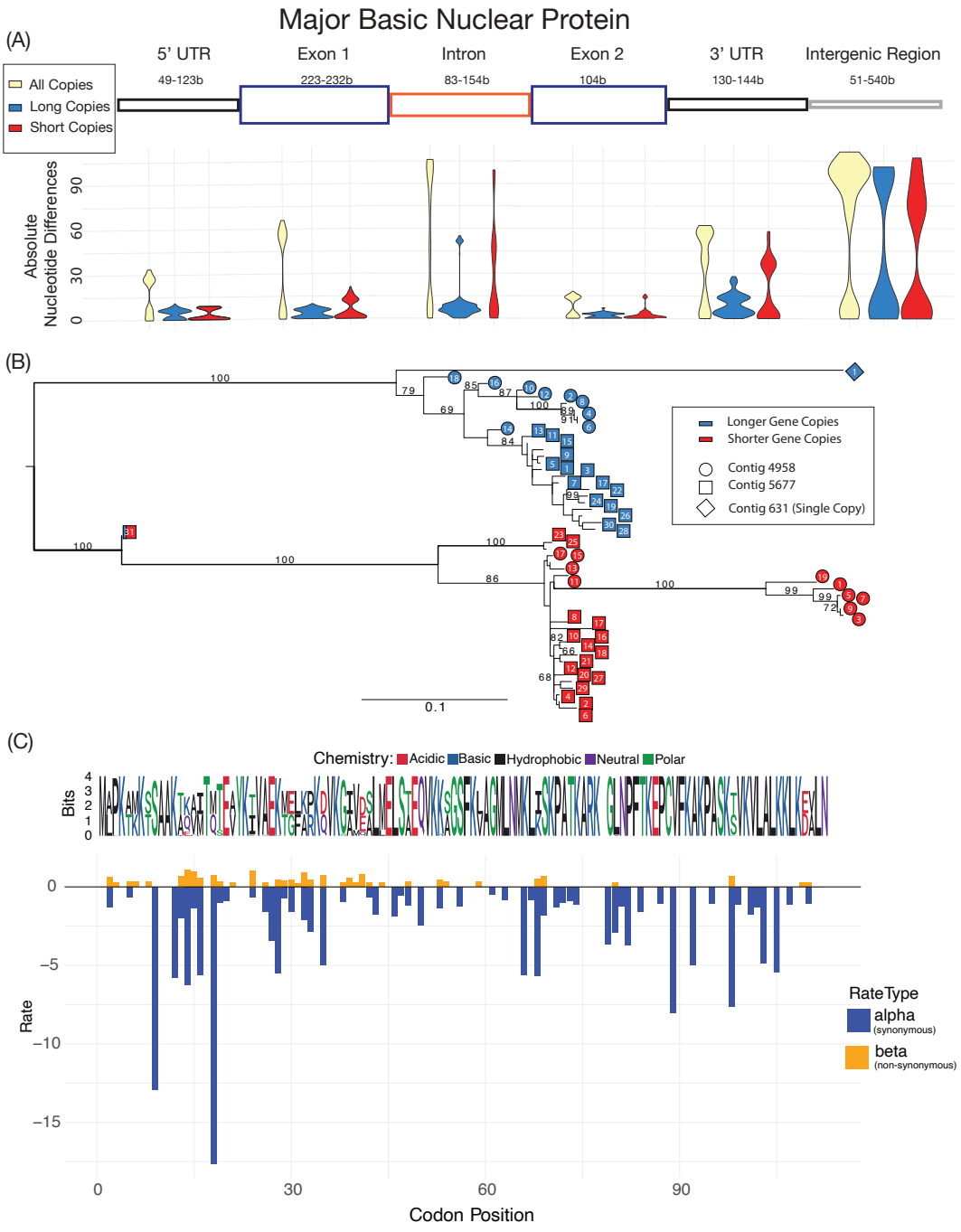


Figure xxx. *Major Basic Nuclear Protein tandem repeats. (A) A schematic of the MBNP gene structure with violin plots of the variation found in each section for comparison of all the copies, and then within the longer and shorter versions. (B) Phylogenetic tree for the sequences of all the copies of MBNP, with shape and color denoting contig and position within the repeat array as*

well as by short vs. long variant, (C) Sequence-logo for all MBNP copies and the site-specific codon analysis performed using the Fixed Effects Likelihood (FEL) method for all MBNP copies.

The long and short variants of the MBNP gene copies can both be found in two genomic locations, labeled contigs 4958 and 5677. Genomic read coverage contained even coverage and had reads spanning the entire gene array. In a CDS sequence alignment, phylogenetic analysis revealed that the sequences were related firstly by gene-variation-type, with long and short variants grouping with other copies of the same variant (Figure xxx-B). Within those variant groupings, sub-groups with copies on one contig are typically more closely related to each other than to the same variant copies on the other contig. Notably, the 31st copy on contig 5677 seems to be a merged version of the variants, with a 5'UTR similar to the long variants, but with other sequence features for the short variant.

In the MBNP gene copies, most sites exhibit $\alpha > \beta$ (synonymous > nonsynonymous rate of mutation), indicating strong purifying selection across the gene, which is consistent with the highly conserved protein sequence and its essential DNA-binding function (Figure xxx-C). The resulting dN/dS ratios are typically less than 1, reaffirming that evolutionary constraints are acting to preserve amino acid integrity. A small number of codons showed elevated β values, suggesting that limited sites may be under relaxed or positive selection, possibly reflecting functional divergence between long and short variants. Notably, no widespread positive selection has been detected, further emphasizing the functional conservation of MBNP across the 52 copies.

Selection on Tandem Genes

The average dN/dS ratios for all of the identified tandem gene families is less than 1, except for the sodium-dependent phosphate transporter genes, which are found in 25 copies across 4 different contigs (Supp. Table x). Even within this tandemly repeated gene group, an average of

9.3% of the sites show diversifying selection while 26.3% are observing purifying selection. About 94% of the tandem gene groups display more sites under purifying selection rather than diversifying selection.

Gene Expression

Tandem genes have higher average expression than single-copy or non-tandem genes, with expression increasing as the copy number increases. The average overall FPKM for identified transcripts is 16.54, with the non-tandem and tandem genes having average FPKM of 7.71 and 434, respectively (Figure xxxi-A). This estimates an average 56X increase in expression in the tandem gene group compared to all other nuclear genes. Expression of the tandem genes increased with copy number, with FPKM increasing by an average of 41.3 per copy with an R^2 of 0.6 and a p-value of 1.74e-15 (Figure xxxi-B). The MBNP gene set was excluded from this analysis because of extreme gene expression, with an FPKM of 6,044, which is 2.5X greater than the next tandem gene set (see inset in top right of Figure xxxi-B).

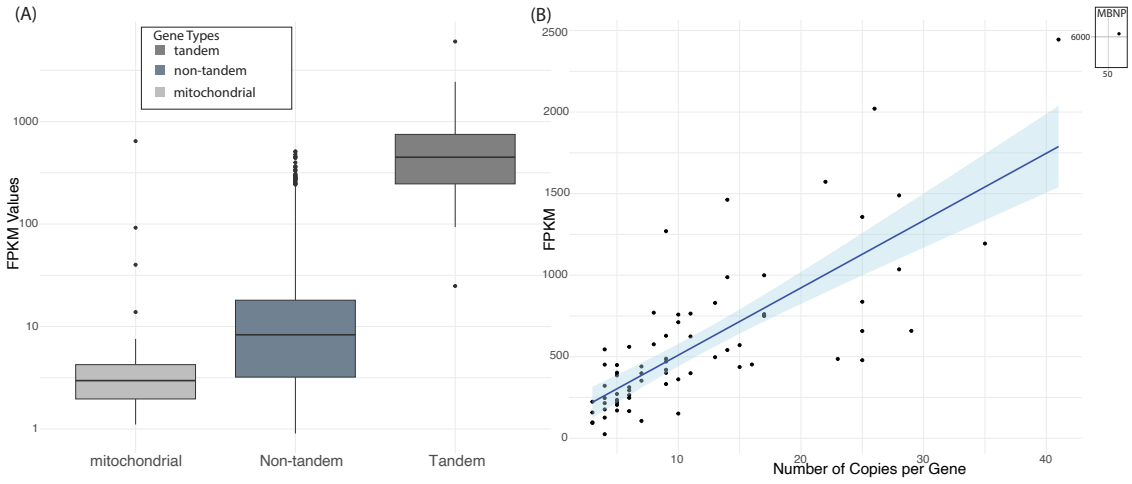


Figure xxxi. *(A) Distribution of FPKM for mitochondrial, non-tandem nuclear, and tandem nuclear genes. (B) Expression of tandem genes on the y-axis compared to their copy number on the x-axis.*

Telomeric Sequences

Telomeric regions, which should correspond to chromosome ends, are present in 23 contigs, with a maximum continuous repeat length of 79,639 bases on a contig 889,787 bases in length (about 9% of the total length) (Supp. Table xi).

Amphidinium carterae **assembly comparisons**

When compared directly for contiguous matches, we found that 5.23% of the present genome aligns to the 2024 Li *et al A. carterae* (unknown strain) genome with 75-100% identity, 26.8% of the assemblies align with 50-75%, 39.5% align with 25-50%, 7.21% align with >0-25%, and 21.29% of the assembly has no alignment (Supp. Figure vii, Supp. Table xiii). The assembly regions with no alignments included 3 contigs from the present study and 4 contigs from the Li *et al* assembly, which are all contigs not scaffolded within the pseudo-chromosomes. The largest continuously aligned region between these two strains is 361,618 bases in length. For context the rDNA regions in both genomes contain identical SSU and LSU regions and both genomes contain rDNA versions one or two bases different across this entire region. This very high identity is also seen for short regions within both assemblies with many regions aligning at 98% identity or higher.

Tandem Genes Across Dinoflagellate References

Despite relatively little overall contiguity between the present and Li *et al* 2024 assemblies, the tandemly repeated genes described in this paper are mostly preserved to a high degree between the two *A. carterae* strains, with very similar or identical arrangements and typically with very high identity nucleotide matches of 98% or better. When directly compared, the average

difference in copy number per gene annotation is very low (<0.003), displaying no consistent bias toward one dataset having more copies than the other (Figure xxxii-A). There is no consistent bias toward one assembly having generally increased or decreased copies as well. For example, the Li *et al* 2024 assembly displays 4 more copies of MBNP compared to the present assembly, and the present assembly contains 4 more copies of actin. This lack of consistent difference suggests these variations are biologically derived between the strains, rather the result of assembly error.

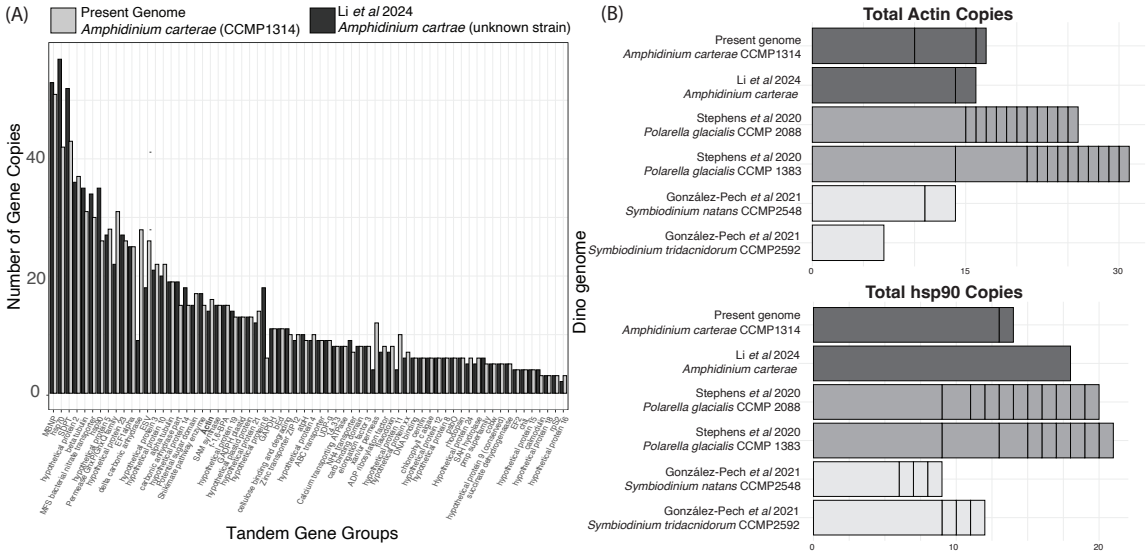


Figure xxxii-A. *Comparison of tandemly repeated gene families in other dinoflagellate references. (A) Tandem gene families found in the present assembly for A. carterae NCMA 1314 compared to Li et al.'s Illumina/PacBio/Hi-C assembly for another strain of A. carterae. (B) Gene copy counts of the conserved and tandemly repeated actin gene (above) and hsp90 (below) across six dinoflagellate genomes, with the bars divided by their distribution per final reported contig/scaffold. SDPT - sodium-dependent phosphate transporter, MSF - Major Facilitator Superfamily (MFS) transporter, f-1,6-BPA - fructose-1,6-biphosphate aldolase, bEd - bacterial Exonuclease/Endonuclease/Phosphatase domain, UDP-g - UDP glucose starch glucosyltransferase*

Tandemly repeated gene groups in the current assembly versus the Li *et al* 2024 assembly are also in highly conserved regions. In the present assembly two tandemly repeated sets of actin genes, which contain 10 copies and 6 copies, correspond to the two major clades previously described with many synonymous substitutions and intron length differences between the clades (Bachvaroff & Place 2008). The longer actin region shares 18,339 bases of sequence with 99% identity with 9 of the 10 genes present in the Li *et al* genome. There are identical intron-exon boundaries, 3' and 5' UTR regions and very high identity across the transcript and intergenic regions. In this 10 gene set, across the entire gene of ~1800 bases including an intron and 1,293 coding bases, there are 6-9 nucleotide differences between the most similar copies across assemblies. Similarly, for the set of 6 tandem repeat actins, the overall identity is 99% over 10,987 bases with one less gene in the Li *et al.* assembly. Thus, the *A. carterae* strain NCMA 1314 has loci of 10 and 6 actin genes while the Li *et al* strain has homologous loci with 9 and 5 gene copies (Supp. Figure viii). Between these two strains the overall picture of tandem gene duplication is very similar.

Finally, the actin and hsp90 genes were used as examples to test if tandem duplication of these conserved and highly expressed genes is also found in four other core dinoflagellate genome assemblies (González-Pech *et al.* 2021; Stephens *et al.* 2020). Comparing the genomes for both strains of *A. carterae*, two *Polarella glacialis* strains, *Symbiodinium natans* and *Symbiodinium tridacnidorum*, there was a range of 7-31 copies of actin across 1-12 scaffolds and a range of 9-27 copies of HSP90 across 4-11 scaffolds (Figure xxxii-B). Overall, the *P. glacialis* strains appeared to have the highest copy numbers for both genes, which aligns with their genomes being over 2X larger than the *A. carterae* strains. The Symbiodiniaceae species have the least number of gene copies for actin and hsp90, with the free-living species having a larger number of actin gene copies, and the symbiotic species having a larger number of hsp90 gene copies. Although the

number of repeats differ between species, and even strains, there appears to be consistent tandem repeat regions for both of these conserved genes in all of the analyzed dinoflagellate references**.**

Discussion

We assembled a near-complete, well-annotated genome for *Amphidinium carterae NCMA 1314*, a free-living basal dinoflagellate known for producing amphidinol toxins. This is one of the first core dinoflagellate genomes sequenced exclusively using long-read ONT, allowing for an un-gapped assembly with potentially less collapse of repetitive regions than typically seen in short-read assemblies (Beedessee *et al.* 2020). Contig N50 for prior dinoflagellate genomes, including *Amphidinium gibbosum*, *Polarella,* and Symbiodiniaceae, ranges from 33kb to 360kb with maximum lengths of 3.4Mb, close to half of the maximum contig length seen in the present *A. carterae* assembly (Supp. Table xii) (Beedessee *et al.* 2020; González-Pech *et al.* 2021; Stephens *et al.* 2020)**.** Here, we were able to assemble one of the longest reported contiguous core dinoflagellate sequences over 7.7Mb. Our genome estimate (~1.2 Gb) aligns with recent findings and is approaching the estimated genome size based on k-mer analysis (Li *et al.* 2024b), transcript mapping, BUSCO scores, and comparison between strains suggesting good completeness. The novel gene annotation presented here relies upon transcriptome-guided evidence, with particular emphasis on the tandemly repeated genes. Fully automated prediction tools may commonly miss tandemly repeated genes or classify them as non-coding repeats, indicating the need for manual curation. Overall, this assembly highlights the usefulness of ONT long-read technology in assembling complex genomes, providing an important resource for functional and comparative genomic studies.

Nuclear Genome

The nuclear genome showcased high levels of repetition, commonly observed in prior dinoflagellate assemblies. The *A. carterae* genome is 55% repetitive, comparable to the 55–70%

repeat content seen in other dinoflagellate assemblies (Allen *et al.* 1975; González-Pech *et al.* 2021; Jaeckisch *et al.* 2011; Li *et al.* 2024b; Stephens *et al.* 2020). We observed widespread tandem gene duplication, a pattern consistent with other species such as *Polarella glacialis* (Stephens *et al.* 2020). Previous analysis of *A. carterae* using PCR-based methods have shown tandem duplications with some intronic regions, and here we observed exon numbers ranging from 1 to 29 in these tandemly repeated copies. Intron-inclusion in the repeated genes suggests DNA based duplication (Bachvaroff & Place 2008). Tandem gene arrays, such as the MBNP gene family, provide more detailed insight into the underlying duplication mechanism. The MBNP genes occur as paired variants—short and long forms—and cluster more closely within their length groups and genomic neighborhood, suggesting localized, DNA-based duplication, possibly via replication slippage (Figure xxx-A) (Viguera 2001)**.** This DNA-based replication slippage concept is further bolstered by the detailed comparison of the two actin loci where there is a single copy difference between strains at each locus and the intron locations are conserved.

Amphidinium carterae reference assemblies are very similar in content and coding regions. Compared to the Li *et al* 2024 assembly of *A. carterae* (unknown strain), assembly statistics seem very similar, including a total assembly length of 1.25 and 1.26 Gb, GC content of 44.1% and 44.8%, repetitive content of 55% and 56%, for the present and Li *et al* genomes, respectively (Supp. Table 6) (Li *et al.* 2024b). Both assemblies also displayed nested genes at rates of 25.2% and 23.9% for the current assembly and the Li *et al.* assembly, respectively. Differences arose due to the novel read lengths observed by ONT, where the contig N50 was 1.15 Mb compared to 820 kb scaffolded to 34.2 Mb for the present genome and the Li *et al* genome, respectively. Overall, the picture is very similar for the tandemly repeated genes in *A. carterae* references. The divergence in contiguous matches aligns with prior dinoflagellate research showing high sequence and structural divergence across different strains. A 2021 study of 15 Symbiodiniaceae species showed that some *Symbiodinium* isolates were comparable to that seen between distinct

genera (González-Pech *et al.* 2021). A 2020 study of *P. glacialis* strains also observed very high similarity in non-repetitive genomic regions but avoided comparisons of the repetitive DNA (~68% of the assemblies) (Stephens *et al.* 2020). This is akin to the very strong alignment we found in the actin gene sets between the *A. carterae* assemblies, displaying high conservation of these coding regions. The differences can somewhat be attributed to repetitive DNA content, which likely drives both structural divergence and alignment issues between conspecific isolates.

Mitochondrial genome

Dinoflagellate mitochondrial genomes remain difficult to characterize due to their atypical organization and expression (Gagat *et al.* 2017; Haro *et al.* 2024; Jackson *et al.* 2012, 2007; Lin *et al.* 2002; Nash *et al.* 2007; Raven & Beardall 2017). Our assembly recovered multiple contigs representing diverse mitochondrial configurations, read depth, and RNA-seq mapping, likely reflecting within-population or even within-cell genomic heterogeneity. The GC content (30%) supports previous findings of an AT-rich *A. carterae* mitogenome (Nash *et al.* 2007). Our results suggest that mitochondrial genome organization is variable, possibly due to recombination and copy number variation as has been previously demonstrated at the mRNA (Chaput *et al.* 2002), genomic (Waller & Jackson 2009), and assembly level (DeMontigny & Bachvaroff 2025). Further research on this variability is needed at the single-cell level to understand this diversity.

Gene Structure

Microexons: Microexons, which have been identified within the present assembly, are involved in a growing field of study as elements which can greatly affect gene products (Ustianenko *et al.* 2017)**.** In the present genome we have identified cases of micro-exons where an exon can be as short as 3 nucleotides, similar to microexons found in *Drosophila,* mice, and chickens (Beachy *et al.* 1985; Cooper & Ordahl 1985; Ustianenko *et al.* 2017)*.* These microexons can have an extreme impact on development and adaptability, as the inclusion/exclusion of the microexon in isoforms

can change the protein structure and, in some cases, the frame of the transcript which can have major impacts on the subsequent product and completely alter the transcriptional response. These small exonic regions can not only hinder common annotation methods, but they may be an area of research that could possibly describe the abnormal transcriptional response observed in dinoflagellates (Roy *et al.* 2018)**.** This area of research into microexons is still in its infancy for dinoflagellates and requires further understanding.

Annotation: The intricate genomic architecture of dinoflagellates poses significant challenges for standard annotation pipelines. Our data identified at least two broad gene structures that are challenging to annotate (1) conserved tandem arrays (Figure xxvii-F) and (2) dispersed, multi-exon genes with long introns and short exons (Figure xxvii-E). The tandem gene sets that were bolstered by manual curation are likely accurate in terms of overall count, arrangement and overall expression pattern. The longer, dispersed multi-exon genes with short exons and long introns most likely remain underestimated. Shorter genes are more easily annotated using both read mapping and other tools. Intuitively, longer exons are readily recovered even with a simple pairwise identity search between transcript and genome using BLAST or exonerate whereas longer genes with short exons are harder to identify and require longer contig sizes to encapsulate the entire transcript on one contiguous genomic sequence. As long-read sequencing technologies become more prevalent, which should expand potentially collapsed tandem repeat regions, these annotation challenges are likely to extend to other dinoflagellate genomes. The insights gained from our study not only enhance the accuracy of gene annotations in dinoflagellates but also contribute to a deeper understanding of their unique genomic features, thereby advancing the broader field of dinoflagellate genomics.

Dosage & Expression: Tandem arrays tended to correlate positively with transcript abundance, suggesting a dosage-based expression mechanism to increase representation in the transcript pool.

This is consistent with prior models in *Drosophila* showing increased expression from repeated genes, where each copy was seen to increase expression up to 5X (Loehlin & Carroll 2016). For dinoflagellates, where constitutive expression under different conditions is common (Lin 2024; Roy *et al.* 2018; Seveno *et al.* 2020), tandem duplication may provide a strategy to modulate transcript abundance. While exon number alone was not significantly correlated with expression, highly expressed genes tended to be structurally simple—shorter in length and with fewer exons (Supp. Figure ix). The short-exon and long-intron gene structures described here may complicate the mapping of RNA-seq data across long gaps. Mapping tools may revert to soft-masking a short, unmapped portion of RNA-seq reads rather than mapping across long introns, which in turn may fragment or shorten predicted transcripts. With these biases in mind, genes with many exons showed more constrained, typically lower expression profiles, followed by nested genes, which are by definition shorter and had higher expression, finally tandem genes were typically highest in expression. These patterns suggest that transcript complexity may modulate transcription along different genomic regions, although the alignment and mapping issues with extremely short exon gene structures could also contribute to lower expression estimates for longer genes.

Further implicated in gene expression and transcript maturation, we identified 180 spliced leader (SL) RNA genes, averaging 59 bases, across 66 contigs. These genes often cluster with tRNAs and 5S rRNA genes, appearing in tandem repeats of 1 to 14 copies. The 22-nucleotide spliced leader sequence is crucial for mRNA maturation in dinoflagellates (Roy *et al.* 2018; Zhang *et al.* 2007b). Similar genomic arrangements have been observed in various dinoflagellate species, with configurations including SL RNA-only tandem repeats, 5S rRNA-only arrays, and mixed SL RNA–5S rRNA clusters, featuring intergenic spacers from 88 bp to over 1.2 kb. This diversity could potentially suggest ongoing genomic duplication and recombination events within each lineage, contributing to the structural complexity of SL RNA genes while preserving their essential function, although this is preliminary. Additionally, in species like *Karenia brevis*, the

association of SL RNA genes with other small nuclear RNA genes, such as U6 snRNA, could indicate coordinated regulation of these RNA elements. The clustering and tandem repetition of SL RNA genes, along with their association with other small RNA genes, may suggest a regulatory mechanism that ensures efficient expression and processing of these vital RNA elements particularly because maturation of every mRNA will consume one transcribed SL.

In parallel, Li *et al.* (2024) uncovered a DNA modification–based regulatory layer in dinoflagellates: 5-hydroxymethyluridine (5-hmU), installed by a ten-eleven translocation (TET) domain and J-binding proteins (JBPs) oxygenase domain-containing enzyme, is selectively enriched over repetitive sequences and transposable elements (TEs) (Li *et al.* 2024b)**.** Crucially, inhibiting 5-hmU formation leads to transcriptional activation of these marked TEs, implying that 5-hmU acts as an epigenetic mark for silencing retrotransposons. These mechanisms are distinct yet potentially complementary. While tandem duplication increases transcript abundance directly via copy number, 5-hmU–mediated silencing of TEs may preserve transcriptome integrity by maintaining repression of potentially disruptive repetitive elements—especially important in genomes organized into massive gene arrays. Moreover, TEs can harbor regulatory sequences or affect chromatin architecture; thus, their silencing may have indirect effects on expression of neighboring genes.

Importantly, these insights into the *A. carterae* genomes highlight that dinoflagellate gene expression is still a rapidly emerging and dynamic area of study. Once thought to lack complex transcriptional regulation, these organisms are now seen to employ both structural (copy-number) and epigenetic (modification-based) strategies, signaling that many mechanistic layers remain to be discovered.

Selection in the Genome: Collectors, not Hoarders

Our analysis of tandemly repeated genes in the dinoflagellate genome showed that the majority of gene families are evolving under purifying selection, with ~94% of tandem groups exhibiting a greater proportion of sites under functional constraint than under diversifying selection. This suggests that, despite the prevalence of tandem duplication in dinoflagellate genomes, most duplicates retain essential functions rather than diverging adaptively, indicating a careful curation of the retained gene copies. Such patterns align with observations from other dinoflagellate genomes, where gene duplication is abundant, but sequence diversification is often limited by the need to maintain stability in key pathways such as photosynthesis, nutrient uptake, and stress response (Lin *et al.* 2015; Shoguchi *et al.* 2013)**.** The dominance of purifying selection across tandem duplicates supports the hypothesis that in dinoflagellates, large-scale gene family expansions primarily serve dosage effects or regulatory fine-tuning rather than rapid adaptive evolution (Stephens *et al.* 2020). This strong functional constraint may reflect the ecological importance of maintaining metabolic and physiological consistency in the face of fluctuating environmental conditions, particularly in nutrient-limited marine systems.

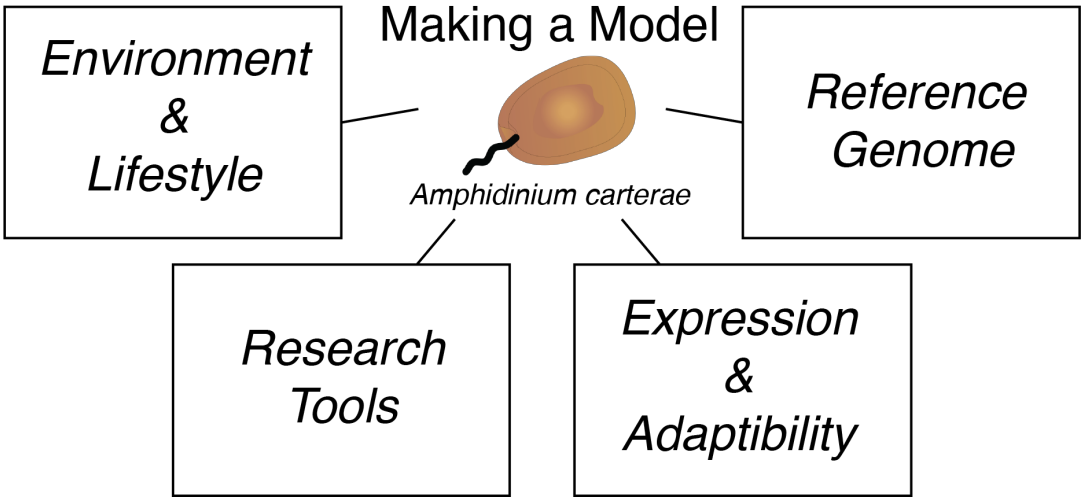
Conclusions

Our assembly reveals diverse gene structures and a striking level of conservation within tandemly repeated genes, suggesting that dinoflagellates actively maintain and regulate these features rather than passively accumulating genomic material. While dinoflagellates can be seen as genomic "hoarders", our findings support a more nuanced view—one in which they are deliberate collectors. Gene copies are preserved for functional and adaptive significance circumstantially correlated with the need for high expression and translation of these genes into protein. The near-complete assembly and comprehensive annotation of tandemly repeated genes presented here offer a critical reference point for future dinoflagellate genome studies, but both assembly and particularly annotation of core dinoflagellate genomes is still at an early stage. Any annotation

method will need to accommodate densely packed tandemly duplicated gene sets, nested genes within long introns, and genes with long introns and relatively short exons. This work lays the groundwork for the development of dinoflagellate-specific gene annotation models, which could streamline annotation pipelines as sequencing technologies continue to scale toward even larger and more complex genomes.

Chapter 6: Discussion

Model: Where do we stand?



Dinoflagellates present a complex system of gene structure and expression that subverts many of the assumptions made for model organisms. In the present study, 4 key areas, 1) exploring Environment & Lifestyle by evaluating culture conditions and characterizing the microbiome, 2) evaluating Research Tools such as genetic knockdowns, 3) analyzing Expression and Adaptability through diel/cell cycle measurements of physical traits and sequencing, and 3) creating a Reference Genome using novel long-read technology, were explored. As described in the introduction, a model organism requires 4 characteristics: 1) Ease of maintenance, 2) Short generation times, 3) Genetic tractability, and 4) Extensive research tools. Ease of maintenance and relatively short generation times were already established. Here we have expanded on the genetic tractability and research tools for *A. carterae* by assembling a nearly complete reference genome and applying novel techniques such as Nanopore Direct RNA sequencing, ribosome profiling, genetic knockdowns and puromycin incorporation in order to better understand the expression dynamics in this species. These steps have increased the support of these key model-organism pillars.

Of the 4 key characteristics necessary to establish a model least developed are research tools, such as protocols for knock-outs and knock-ins to completely remove or introduce specific gene's effects. Although we have produced important milestones in establishing *A. carterae* as a model dinoflagellate, more steps are needed in order to use this species as a model in typical genetic experiments. This includes the ability to create and screen mutants in order to unravel the functions of the "dark genes" described in dinoflagellate systems (Stephens et al. 2018). In typical models, mutants can be produced via random mutagenesis (via chemicals, radiation, and transposons) or targeted mutagenesis (site-directed mutagenesis) and screening of mutants for specific markers or changes in phenotype can indicate the function of the mutated region in the genome. The genetic knockdown established in Chapter 3 is a path towards beginning to observe phenotypic differences, but not as specific and complete as knock-out methods. *Amphidinium carterae* is typically grown in a planktonic population as well, making screening difficult without a sorting cytometer, compared with plating on semisolid media. Culturing dinoflagellates as spheroplasts may be an avenue for screening that would allow visualization of phenotypic differences, although this method does greatly affect cell shape and motility. In a study by Kwok et al (2007), PEG-subjected spheroplast *Crypotodinium cohnii* cells were able to grow, generate cell walls, and go through cellular division (Chan et al. 2019; Kwok et al. 2007). Production of spheroplast dinoflagellate cultures may be the next step in mutant screening. In future steps, plating of *A. carterae* on PEG-containing agar media could be followed by the introduction of a mutagen to create new phenotypes, such as transposons to "break" genes. Interesting phenotypes could be sequenced and compared to the present genome assembly in order to begin to elucidate "dark gene" functions.

Dinoflagellate Expression

Gene complexity and tandem copy number seem to directly impact transcription in dinoflagellates, with lower complexity genes and higher copy numbers allowing for greater

transcriptional activity. In eukaryotes, transcription and translation occurs separately organizationally and temporally, leading to multiple checkpoints of regulation. In dinoflagellates, the nucleus lacks conventional histones to regulate availability of genes for transcription, with some research theorizing that the consistently condensed crystalline chromosome structures may impact transcriptional availability, as well as the use of modified bases (Marinov *et al.* 2024). Because of this distinction in dinoflagellates, the checkpoints for expression likely differ in some respects from conventional eukaryotic mechanisms. In order to pinpoint these points of regulation, the present dissertation aimed to observe differences in the DNA, RNA, and protein pools of *A. carterae* through analysis of the genome, RNA levels, active translation (RPFs), and protein abundance (puromycin incorporation). Transcription of genes appeared to be relatively consistent, even across temporal differences, alluding to almost constitutive transcription. The ribosome profiling data displayed a pattern of greater differential expression, with key transcripts being actively translated in accordance with the time of day and cell cycle. Lastly, translation rate measurements showed an increase in translation over the day and a decrease in the night, underscoring the availability and activity of the ribosomes as they selectively translate from the RNA pool. These observations indicate that key regulation checkpoints may be more crucial at the genomic and translation levels and less emphasized in transcription compared to other eukaryotic systems. Mechanisms affecting gene structure, such as intronization (through horizontal gene transfer, transposition, and DNA repair pathways), as well as gene duplication (through DNA slippage) affect the availability and transcription rate of genes (Koonin 2006; Viguera 2001). This is then followed by constitutive, yet unequal, transcription across the genome, with factors such as gene complexity and copy number having a greater effect on the availability of transcripts from the RNA pool. After export and processing, the transcript is shuttled to the available ribosome to be translated – in this step, there are likely factors such as eIF4Es that regulate the movement of the transcripts as well as the availability and activity of the

ribosomes (Jagus *et al.* 2012; Ross-Kaschitza & Altmann 2020). The observations in this study create a global map of these key regulation steps, and further research should explore these areas for specific mechanisms in order to better understand how they function.

Dinoflagellate Genomics

In the previous chapters, we characterized gene structures that are common in *A. carterae*, but less common in many model systems. This is initially pronounced in the sheer volume of DNA found in dinoflagellate cells, with our *A. carterae* genome having 1.2 Gb. Dinoflagellates are generally observed to be haploid, only briefly entering a diploid state during sexual reproduction, so the present *A. carterae* genome would estimate the DNA content of an average cell to be 1.25 pg, compared to 6.5 pg in *Homo sapiens* (diploid), 0.44 pg in *Arabidopsis thaliana* (diploid), or 0.12 pg in *Chlamydomonas reinhardtii* (diploid) (Chaux-Jukic *et al.* 2021; Lin 2011; Schmuths 2004; Sun *et al.* 2018). Previous research into dinoflagellate genome sizes may have overestimated the DNA content of dinoflagellate cells, which could be due to their unusual genome structures and liquid crystalline state (Chow *et al.* 2010; Wong 2019). One study by LaJeunesse et al (2005) measured DNA content in a variety of dinoflagellate species using DAPI staining compared to chicken red blood cells and *Arabidopsis thaliana* and *Nicotiana tabacum* controls, although these controls have more conventional nuclei structures (LaJeunesse *et al.* 2005). LaJeunesse *et al* approximated *A. carterae* DNA content to be 5.9pg. Assuming *A. carterae* is haploid for most of its life, similar to other dinoflagellates, this is over a 4X overestimation. There is a chance that dinoflagellates may have observed whole genome duplications (WGD), which would cause extreme genome expansion (Lin 2024). In this case it is possible that the present assembled genome may be one copy of a 4X duplication within the cells, but evidence of WGD duplication in dinoflagellate cells is mixed. Although the amount of DNA in dinoflagellate cells is not larger than some known organisms, the novelty lies in the DNA content in a single-celled algal organism. Although human cells contain higher amount of DNA

than *A. carterae*, comparing the content the plant model, *Arabidopsis*, and the algal model, *Chlamydomonas*, shows a 3X to 10X amount, respectively.

Unique gene structures in the present assembly are also more expansive than expected. The genes characterized by their very short (in some cases 'micro') exons and long introns in the *A. carterae* assembly appear to be ubiquitous across the genome, but similar gene structures are found in humans, such as in the ROBO2 gene, which contains an intron spanning over 1.16 Mb (Piovesan *et al.* 2019). There is also a characterized "mega-intron" found in *Drosophila* in the dynein gene DhDhc7(Y) that actually contributes to the structure of a giant thread loop and takes over 3 days to transcribe (Reugels *et al.* 2000). Microexons are also somewhat common in eukaryotic organisms, with the human genome containing about 13K and the mouse genome containing about 3K (Li *et al.* 2015; Parada *et al.* 2021; Ustianenko *et al.* 2017). In vertebrates, average exon length were found to be 188 bp, and in humans genes have an average of 8.8 exons with 80% being < 200 bp (Li *et al.* 2017; Sakharkar *et al.* 2004). In the present genome, exons count per gene were similar to those found in complex eukaryotes, with there being an average of 7.9 exons per gene and an average exon length of 123 bases. Other models such as *Arabidopsis* have average exon lengths of 240bp with 5.4 exons per gene, and *Chlamydomonas* has exon length averages of 240bp with 7.3 exons per gene, overall having longer exons and less exons-per-gene compared to *A. carterae* (Baier *et al.* 2018; Swarbreck *et al.* 2007; Town *et al.* 2006).

Repetitive content is also common in eukaryotes, with genomes containing 25-50%repetitive regions (Liao *et al.* 2023). Genomes for *Drosphila* and *Arabidopsis*, two widely used model organisms, contain repetitive DNA content of 35% and 14-27%, respectively, which is less extreme compared to mammals, and Chlamydomonas contains up to 22% repetitive regions. The present *A. carterae* genome assembly sits closer to the extent seen in mammals, such as humans with >50% repetitive regions (Lander *et al.* 2001).

Tandem gene arrays are also common in eukaryotes and are often caused by unequal crossing over during meiosis and replication slippage (Birchler & Yang 2022). The effect these tandem gene arrays can have on subsequent expression is measurable; for example, in a study by Loehlin and Carroll (2016), they showed that addition of duplicated reporter genes can have a 2.3-5.1-fold increase in expression levels per copy number. In vertebrates, tandem gene arrays account for approximately 14% of all genes, typically in the same transcription orientation, although the majority of these tandems only have 2 copies (Pan & Zhang 2008). Of these tandemly repeated genes in vertebrates, 18% to 40.3% have greater than 2 copies (about 2%), which is similar to the 2% found in the present *A. carterae* assembly. The present assembly also displayed high copy count numbers, such as MBNP which had a maximum of 32 copies over one contig. These numbers are high even compared to other eukaryotes, such as in *Arabidopsis* which has a maximum tandemly duplicated copy number of 23 (Simillion *et al.* 2002).

Genome, Expression, and Environment

Links between constitutive transcription and ecological outcomes like bloom density are still being explored, but this lifestyle may allow for dinoflagellates to actively respond to favorable conditions by bypassing the delays associated with transcriptional regulation. This constant metabolic readiness may underlie their tendency to form rapid and intense blooms. In aquatic environments, microbial responses depend heavily on the timescale of environmental changes versus the microbial response. These responses include gene regulation, nutrient uptake, and growth rates. If an environment shifts slower than the microbial response, then microbes are able to respond in a timely manner to those changes, but if environmental factors shift faster than microbial response rates, this can leave the microbial population in flux (Nguyen *et al.* 2021). In many aquatic environments, microbial populations display synchronized gene expression, where different populations display coordinated transcriptional cycles in response to environmental cues, although the genes expressed are variable (Ottesen *et al.* 2013). In environments that

accumulate dinoflagellate blooms, major shifts in gene expression are typically driven by species succession (Metegnier *et al.* 2020). Research has shown that dinoflagellates seem to express a 'core' gene expression profile, where the RNA pool remains static under different conditions (Moustafa *et al.* 2010). Constitutive transcription in dinoflagellates may also allow for robustness in shifting environments, which is apparent as dinoflagellate blooms can form later than other species under stratified, nutrient-limited conditions. This does come with a tradeoff of dinoflagellate populations generally growing slower than other aquatic microbial populations, likely due to their genome size and complexity, cell size, and perhaps due to energy expenditure on constitutive transcription (Ollison *et al.* 2023; Peltomaa *et al.* 2019). In bloom-forming environments, this typically means that faster-growing populations of aquatic microbes dominate earlier during favorable conditions, with dinoflagellate blooms occurring later after this initial response. The prevalence of dinoflagellates and their robustness in aquatic environments is still in need of further investigation. Dynamics of bloom formations are well characterized, but the mechanisms allowing for dinoflagellate responses is not fully understood. Therefore, understanding how dinoflagellate regulate these adaptations to their environment is crucial for understanding growth dynamics and accumulation of harmful algal blooms.

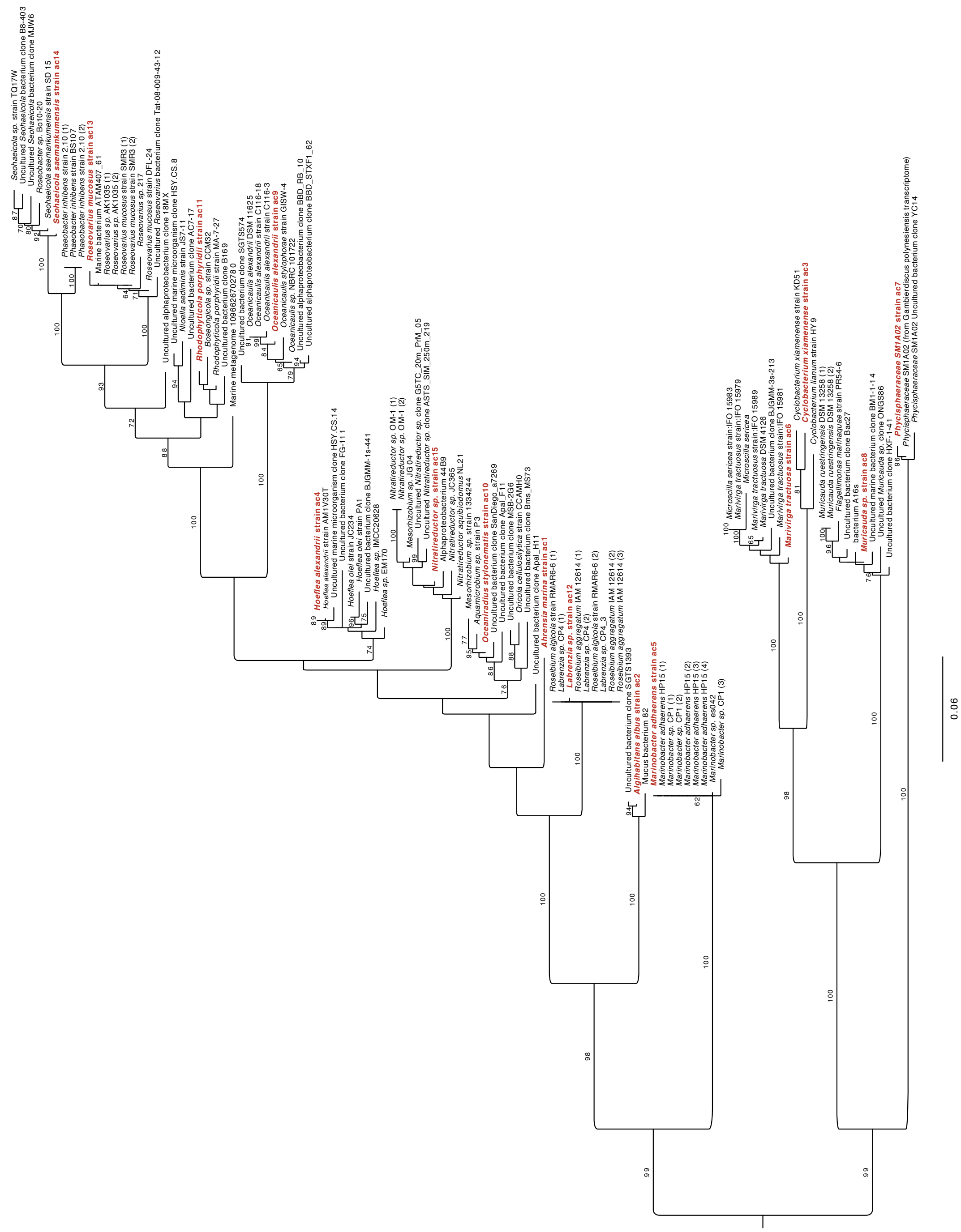
Dinoflagellates are not solitary actors in marine ecosystems but engage intimately with a dynamic microbiome that impacts their physiology and ecological success. Studies of *Prorocentrum* species reveal that their core bacteriome—consisting primarily of Alphaproteobacteria, Gammaproteobacteria, and Bacteroidia—likely supports essential functions such as nutrient exchange, stress mitigation, and secondary metabolite production within the phycosphere (Martínez-Mercado *et al.* 2024). In the benthic, toxin-producing genus *Gambierdiscus*, associated bacterial communities appear responsive to nitrogen availability and correlate with shifts in toxin production, underscoring their active role in bloom dynamics and toxicity regulation (Wu *et al.* 2022). In the present studies, we were able to characterize the microbial population associated

with our cultures *A. carterae* species, and pathways likely leading to close-knit interactions. At least some of these bacterial species are derived from the environment, with common species such as *Roseovarius*, *Seohaeicola*, and SM1A02 being found in other dinoflagellate communities as well (Biebl *et al.* 2005; Miller & Belas 2004; Tizabi *et al.* 2025; Yoon *et al.* 2009). *Amphidinium carterae* is the species of interest but understanding its role and connections to the microbiome also allow us to parse what aspects of the environment and microbial response are from the dinoflagellate versus its microbiome.

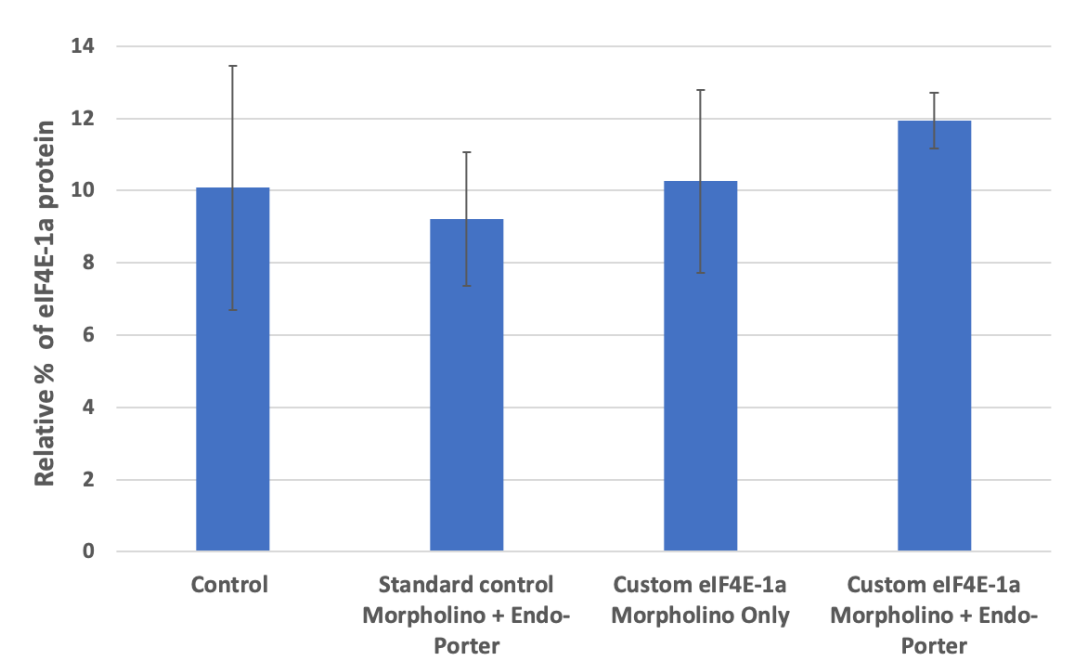
Conclusions

The groundwork for establishing *A. carterae* as a model for dinoflagellates has been described in the previous chapters, with analysis into species expression and genomics. As research into this area progresses, the ability to conduct genomic experiments on *A. carterae* will allow for a greater understanding of the dinoflagellate genome and its associated expression. Currently we have a basic understanding of the rate-limiting steps in expression, from gene-copy-number to ribosome activity, and have observed de-emphasis on transcriptional control in regard to the dinoflagellate central dogma. The next order or understanding will be a systematic way of altering and screening dinoflagellate cells in order to elucidate the function of genes and pathways present in the genome. Understanding of the functions of dinoflagellate genes will aid in our understanding of their ecological role and bloom dynamics, which will allow for greater management responses to environmental and economic disasters.

Appendices

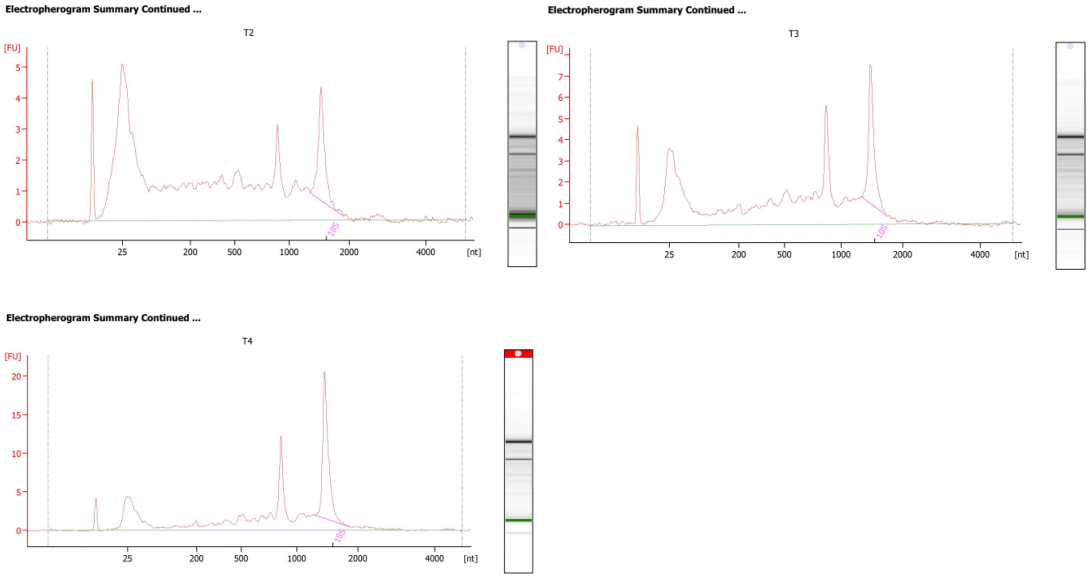


Supplemental Figure i. *Phylogenetic tree including bacterial 16S sequences found in the A. carterae microbiome. Tree was generated using SILVA ACT Service using RaxML, the GTR model, and the Gamma rate model, with a minimum sequencing identity of 0.95 and including 10 neighbors.*

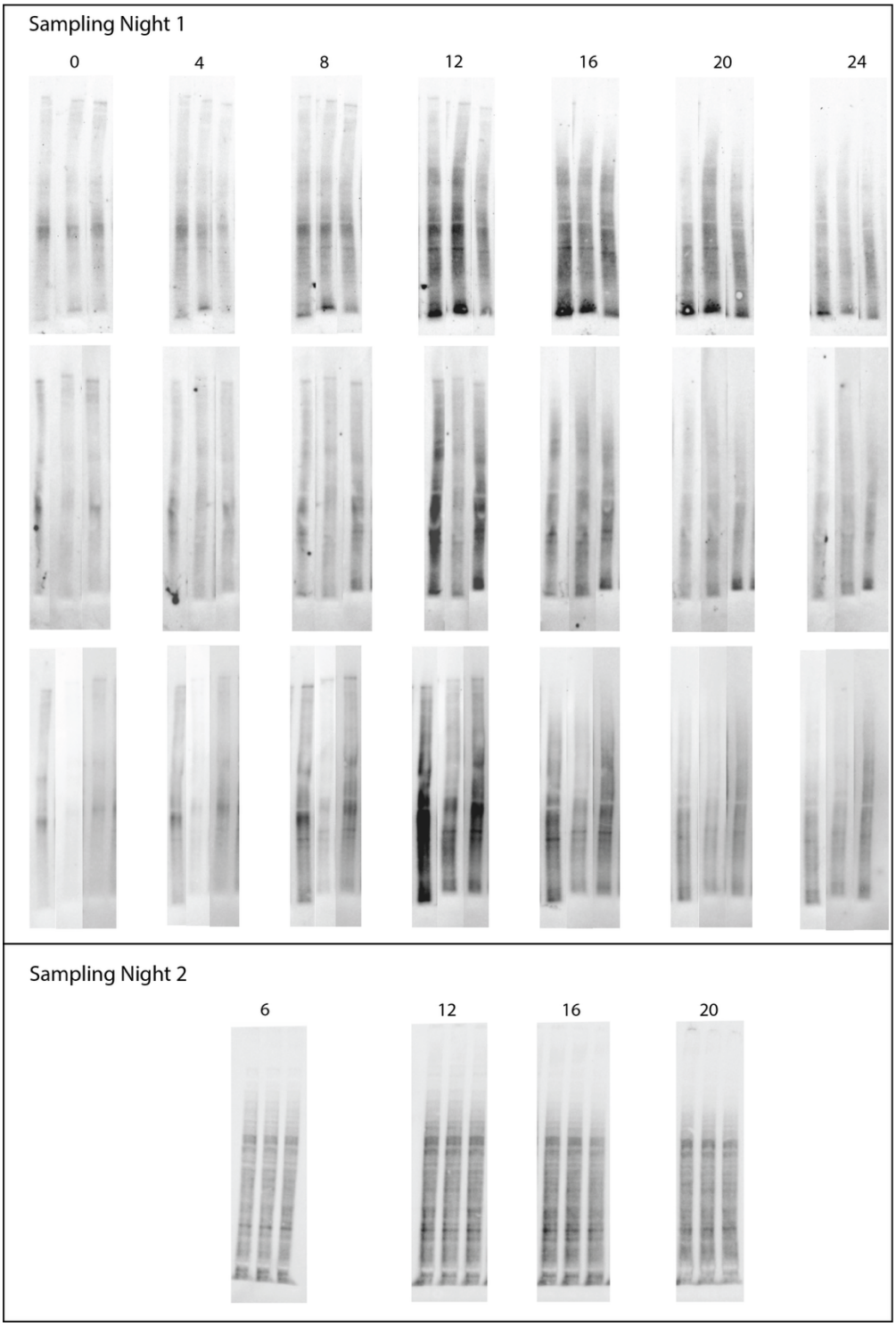


Supplemental Figure ii. *Western blot analyses for eIF4E-1a concentrations within control and treated cells at 96 hours post-MO [10* ***µM****] addition (N=3). Protein loading and relative expression levels was verified by probing with anti-eIF4E-1a mouse monoclonal and HRP-conjugated anti-mouse IgG and compared to total protein volumes. The effect of the custom translation-blocking morpholinos show no effect after 96 hours of treatment. Statistical analysis was done with t-tests using pooled standard deviations.*

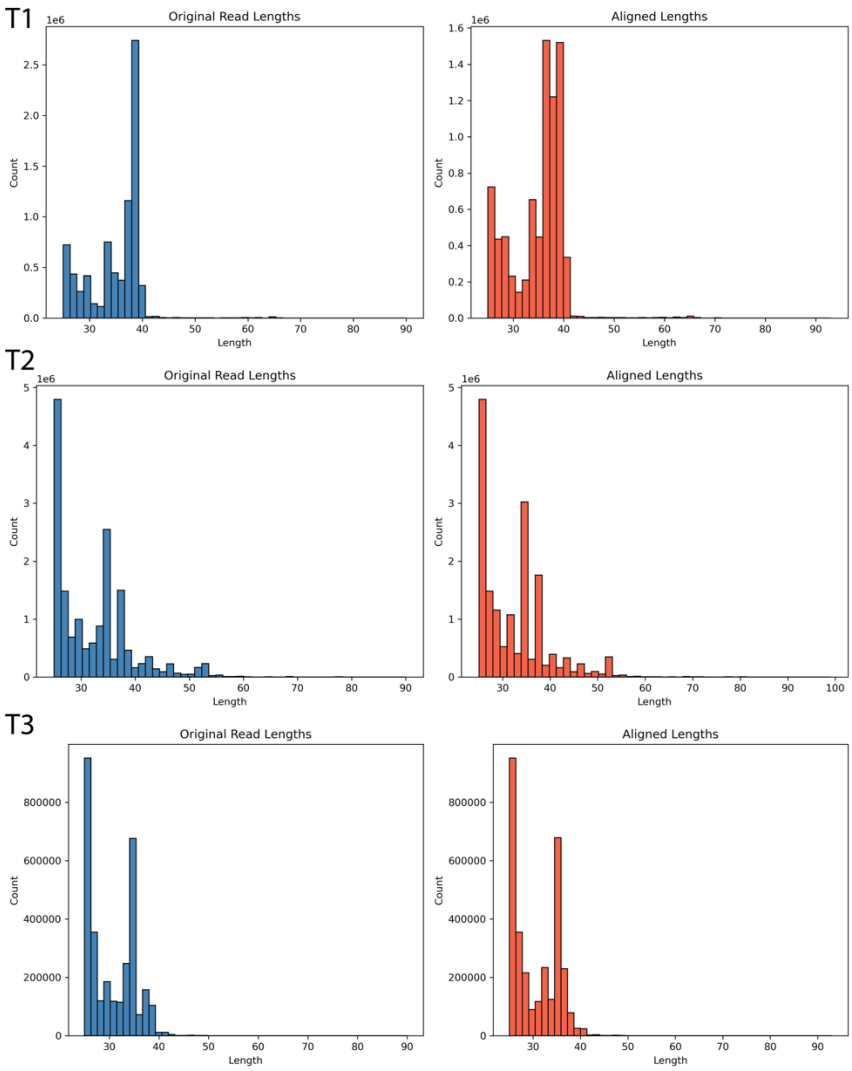
No significant difference between any samples found based on a p-value of < 0.05.



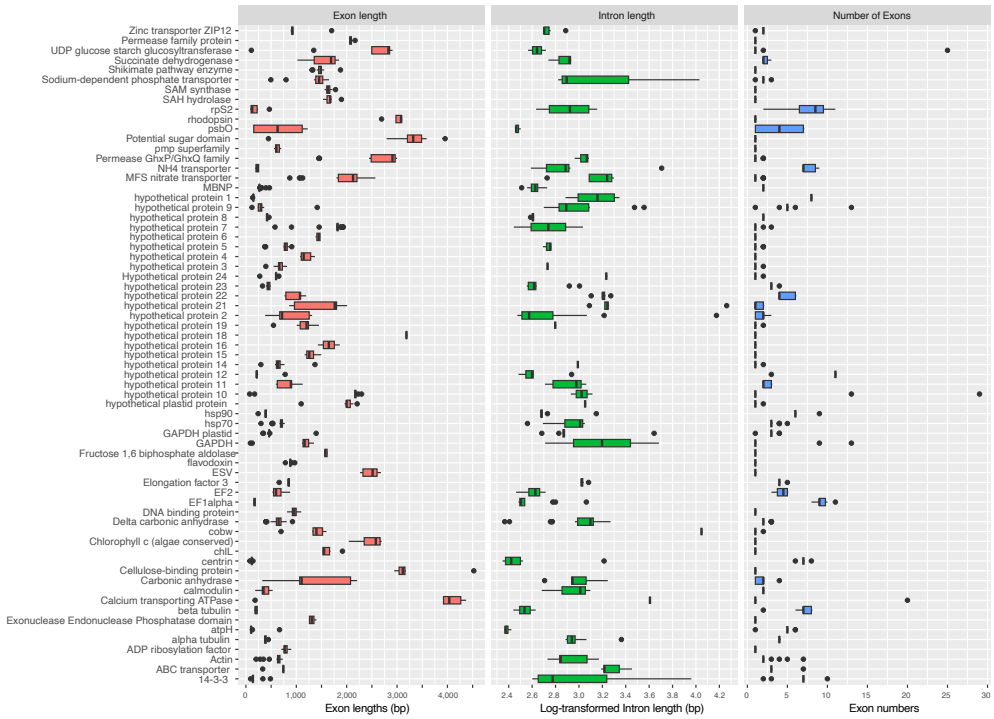
Supplemental Figure iii*. BioAnalyzer results for extracted and isolated RNA before poly-A selection and preparation for Nanopore Direct RNA sequencing*


Sampling Night 1
0
4
8
12
16
20
24
Sampling Night 2
6
12
16
20

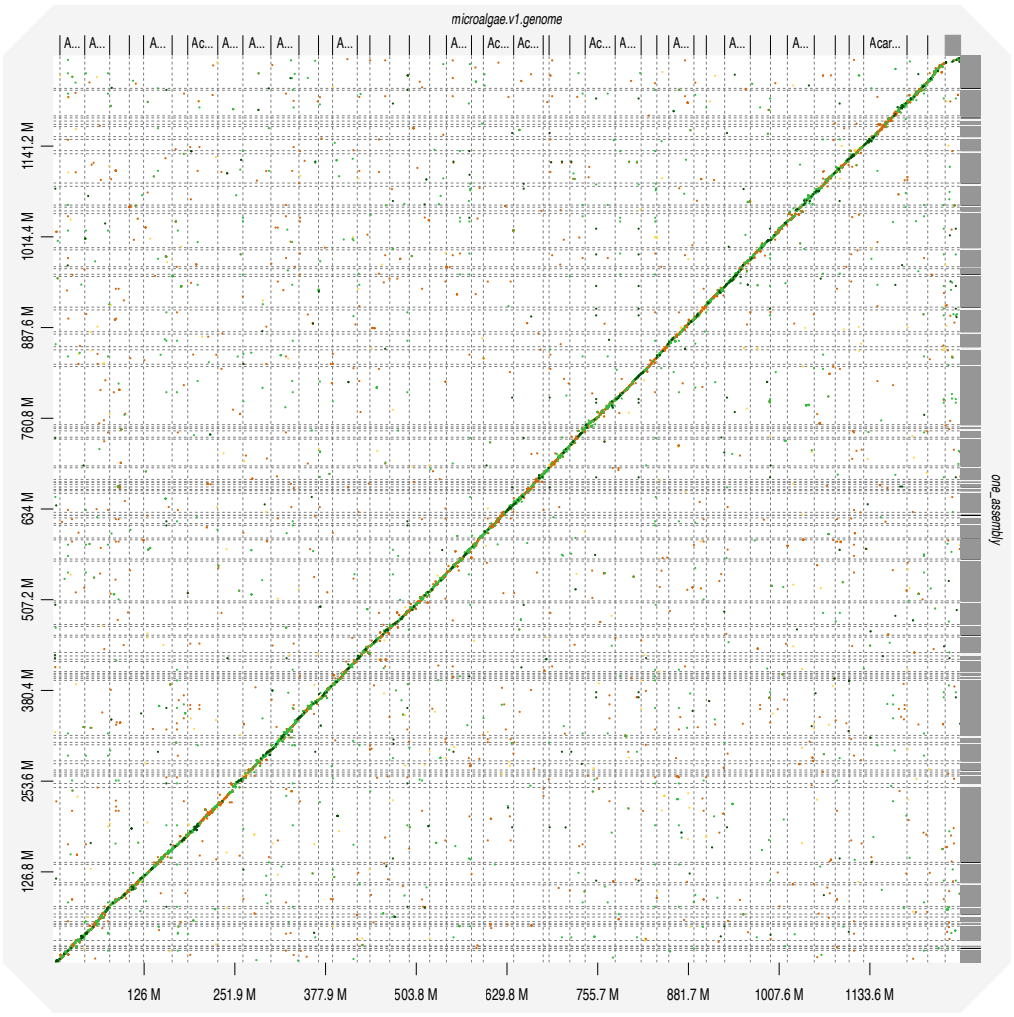
Supplemental Figure iv. *Western blots of A. carterae cell lysates over the diel cycle probed with monoclonal mouse anti-puromycin antibody and a secondary goat anti-mouse HRP monoclonal antibody.*



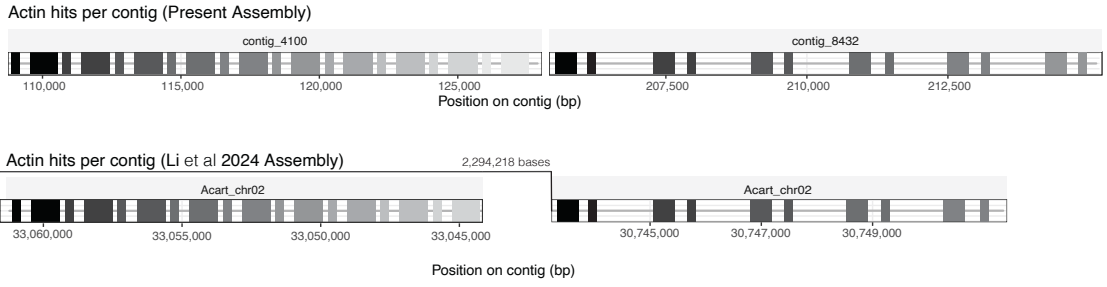
Supplemental Figure v*. Distribution of Ribo-Seq read mapping.*



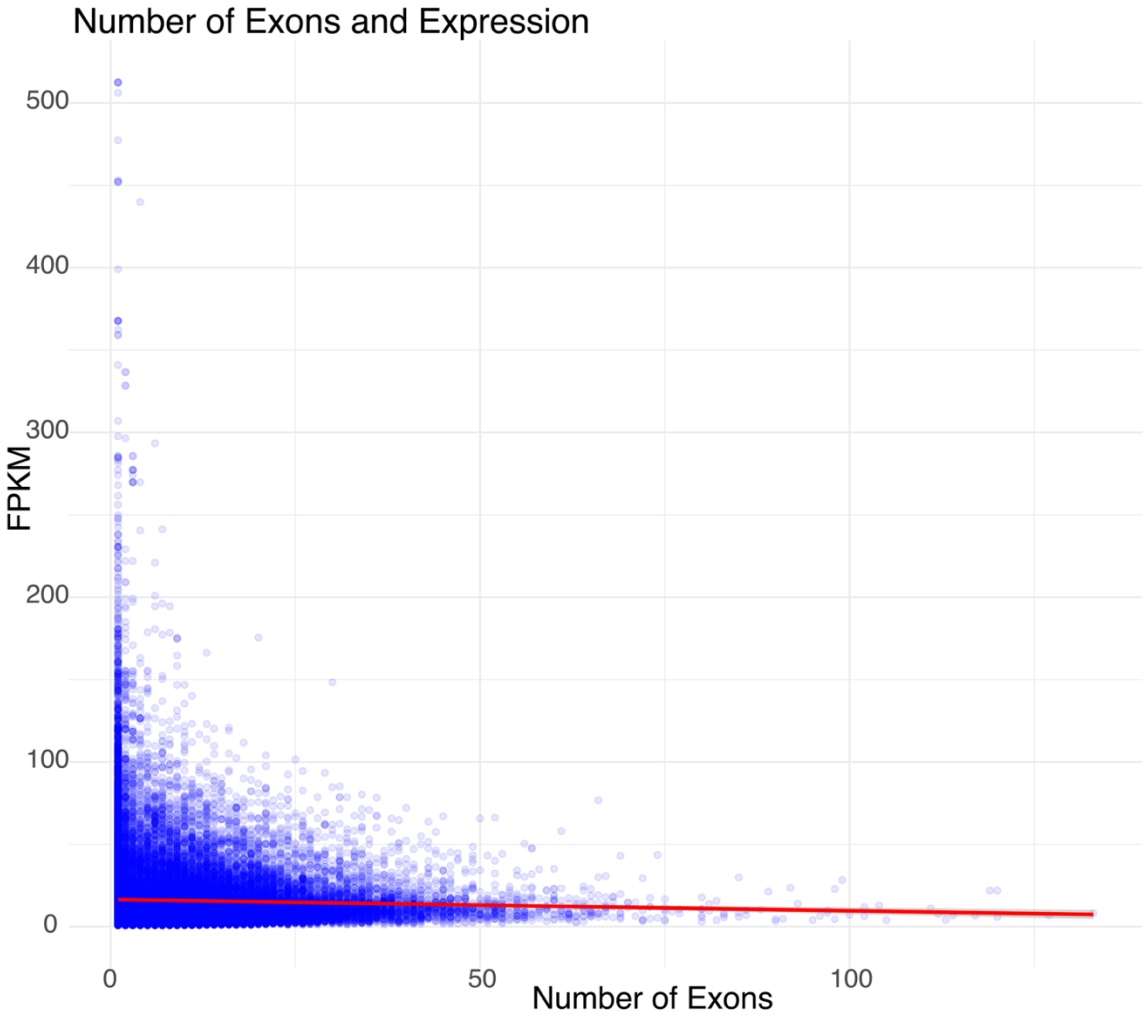
Supplemental Figure vi. Tandem gene structure characteristics. Left: exon lengths, Center: intron lengths (log transformed), and Right: exon number.



Supplemental Figure vii. *D-Genies mapping of the present assembly to the Li et al 2024 assembly of another strain of A. carterae*



Supplemental Figure viii. *Map of tandemly repeated actin gene copies in the present assembly (top) and the Li et al 2024 assembly (bottom).*



Supplemental Figure ix. *Number of exons in single-copy genes and their associated FPKM.*

Bibliography

Adachi, H., De Zoysa, M.D. & Yu, Y.-T. (2019) Post-transcriptional pseudouridylation in mRNA as well as in some major types of noncoding RNAs. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1862**, 230–239

Adolf, J.E., Stoecker, D.K. & Harding, L.W. (2006) The balance of autotrophy and heterotrophy during mixotrophic growth of *Karlodinium micrum* (Dinophyceae). *Journal of Plankton Research* **28**, 737–751

Akbar, M.A., Ahmad, A., Usup, G., *et al.* (2018a) Current knowledge and recent advances in marine dinoflagellate transcriptomic research. *Journal of Marine Science and Engineering* **6**, 1–16

Akbar, M.A., Ahmad, A., Usup, G., *et al.* (2018b) RNA-Seq as an Emerging Tool for Marine Dinoflagellate Transcriptome Analysis: Process and Challenges . *Processes* **6**

Alavi, M., Miller, T., Erlandson, K., *et al.* (2001) Bacterial community associated with *Pfiesteria*-like dinoflagellate cultures. *Environmental Microbiology* **3**, 380–396

Albinsson, M.E., Negri, A.P., Blackburn, S.I., *et al.* (2014) Bacterial Community Affects Toxin Production by *Gymnodinium catenatum*. *PLoS ONE* **9**, e104623

Allen, J.R., Roberts, T.M., Loeblich, A.R., *et al.* (1975) Characterization of the DNA from the dinoflagellate *Crypthecodinium cohnii* and implications for nuclear organization. *Cell* **6**, 161–169

Ambrosino, L., Riccardi, A., Welling, M.S., *et al.* (2024) Comparative Transcriptomics to Identify RNA Writers and Erasers in Microalgae. *International Journal of Molecular Sciences* **25**, 8005

Aranda, M., Li, Y., Liew, Y.J., *et al.* (2016) Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. *Scientific Reports* **6**, 39734

Babaei, S., Bhalla, P.L. & Singh, M.B. (2024) Identifying long non-coding RNAs involved in heat stress response during wheat pollen development. *Frontiers in Plant Science* **15**

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**, e0212912

Bachvaroff, T.R., Gornik, S.G., Concepcion, G.T., *et al.* (2014) Dinoflagellate phylogeny revisited: Using ribosomal proteins to resolve deep branching dinoflagellate clades. *Molecular Phylogenetics and Evolution* **70**, 314–322

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**

Bachvaroff, T.R., Sanchez-Puerta, M.V. & Delwiche, C.F. (2006) Rate Variation as a Function of Gene Origin in Plastid-Derived Genes of Peridinin-Containing Dinoflagellates. *Journal of Molecular Evolution* **62**, 42–52

Baier, T., Wichmann, J., Kruse, O., *et al.* (2018) Intron-containing algal transgenes mediate efficient recombinant gene expression in the green microalga *Chlamydomonas reinhardtii*. *Nucleic Acids Research* **46**, 6909–6919

Barlow, S.B. & Triemer, R.E. (1988) The mitotic apparatus in the dinoflagellate *Amphidinium carterae*. *Protoplasma* **145**, 16–26

Batool, A., Aashaq, S. & Andrabi, K.I. (2019) Eukaryotic initiation factor 4E (eIF4E): A recap of the cap-binding protein. *Journal of Cellular Biochemistry* **120**, 14201–14212

Baumgarten, S., Bayer, T., Aranda, M., *et al.* (2013) Integrating microRNA and mRNA expression profiling in *Symbiodinium microadriaticum*, a dinoflagellate symbiont of reef-building corals. *BMC Genomics* **14**

Beachy, P.A., Helfand, S.L. & Hogness, D.S. (1985) Segmental distribution of bithorax complex proteins during *Drosophila* development. *Nature* **313**, 545–551

Beauchemin, M., Roy, S., Daoust, P., *et al.* (2012) Dinoflagellate tandem array gene transcripts are highly conserved and not polycistronic. *Proceedings of the National Academy of Sciences* **109**, 15793–15798
Beedessee, G., Kubota, T., Arimoto, A., *et al.* (2020) Integrated omics unveil the secondary metabolic landscape of a basal dinoflagellate. *BMC Biology* **18**, 139
Berges, J.A., Franklin, D.J. & Harrison, P.J. (2001) Evolution of an Artificial Seawater Medium: Improvements in Enriched Seawater, Artificial Water Over the Last Two Decades. *Journal of Phycology* **37**, 1138–1145
Bertile, F., Matallana-Surget, S., Tholey, A., *et al.* (2023) Diversifying the concept of model organisms in the age of -omics. *Communications Biology* **6**, 1062
Biebl, H., Allgaier, M., Lünsdorf, H., *et al.* (2005) *Roseovarius mucosus* sp. nov., a member of the Roseobacter clade with trace amounts of bacteriochlorophyll a. *International Journal of Systematic and Evolutionary Microbiology* **55**, 2377–2383
Biebl, H., Pukall, R., Lünsdorf, H., *et al.* (2007) Description of *Labrenzia alexandrii* gen. nov., sp. nov., a novel alphaproteobacterium containing bacteriochlorophyll a, and a proposal for reclassification. *International Journal of Systematic and Evolutionary Microbiology* **57**, 1095–1107
Biémont, C. (2010) A Brief History of the Status of Transposable Elements: From Junk DNA to Major Players in Evolution. *Genetics* **186**, 1085–1093
Birchler, J.A. & Yang, H. (2022) The multiple fates of gene duplications: Deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. *The Plant Cell* **34**, 2466–2474
Bolch, C.J.S., Bejoy, T.A. & Green, D.H. (2017) Bacterial Associates Modify Growth Dynamics of the Dinoflagellate *Gymnodinium catenatum*. *Frontiers in Microbiology* **8**
Bolch, C.J.S., Subramanian, T.A. & Green, D.H. (2011) The Toxic Dinoflagellate *Gymnodinium catenatum* Require Marine Bacteria for Growth. *Journal of Phycology* **47**, 1009–1022
Bowazolo, C. & Morse, D. (2023) Ribosome profiling in the Symbiodiniacean dinoflagellate *Fugacium kawagutii* shows coordinated protein synthesis of enzymes in different pathways at different times of day. *Molecular Microbiology* **120**, 462–471
Bowazolo, C., Song, B., Dorion, S., *et al.* (2022) Orchestrated translation specializes dinoflagellate metabolism three times per day. *Proceedings of the National Academy of Sciences* **119**
Brayton, K.A., Lau, A.O.T., Herndon, D.R., *et al.* (2007) Genome Sequence of *Babesia bovis* and Comparative Analysis of Apicomplexan Hemoprotozoa. *PLoS Pathogens* **3**, e148
Bui, Q.T.N., Pradhan, B., Kim, H.-S., *et al.* (2024) Environmental Factors Modulate Saxitoxins (STXs) Production in Toxic Dinoflagellate *Alexandrium*: An Updated Review of STXs and Synthesis Gene Aspects. *Toxins* **16**, 210
Burki, F., Imanian, B., Hehenberger, E., *et al.* (2014) Endosymbiotic Gene Transfer in Tertiary Plastid-Containing Dinoflagellates. *Eukaryotic Cell* **13**, 246–255
Cabanettes, F. & Klopp, C. (2018) D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* **6**, e4958
Cantarel, B.L., Korf, I., Robb, S.M.C., *et al.* (2008) MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* **18**, 188–196
Cenci, U., Qiu, H., Pillonel, T., *et al.* (2018) Host-pathogen biotic interactions shaped vitamin K metabolism in Archaeplastida. *Scientific Reports* **8**, 15243
Chan, P.P., Lin, B.Y., Mak, A.J., *et al.* (2021) tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Research* **49**, 9077–9096
Chan, W.S., Kwok, A.C.M. & Wong, J.T.Y. (2019) Knockdown of dinoflagellate cellulose synthase CesA1 resulted in malformed intracellular cellulosic thecal plates and severely impeded cyst-to-swarmer transition. *Frontiers in Microbiology* **10**, 1–14

Chaput, H., Wang, Y. & Morse, D. (2002) Polyadenylated Transcripts Containing Random Gene Fragments are Expressed in Dinoflagellate Mitochondria. *Protist* **153**, 111–122

Chaux-Jukic, F., O'Donnell, S., Craig, R.J., *et al.* (2021) Architecture and evolution of subtelomeres in the unicellular green alga *Chlamydomonas reinhardtii*. *Nucleic Acids Research* **49**, 7571–7587

Chen, T., Chen, X., Sun, H., *et al.* (2024) Unveiling the responses of *Alexandrium pacificum* to phosphorus utilization by physiological and transcriptomic analysis. *Science of The Total Environment* **911**, 168759

Chen, X., Yuan, Y., Zhou, F., *et al.* (2025) RNA m5C modification: from physiology to pathology and its biological significance. *Frontiers in Immunology* **16**

Chen, Y., Dougan, K.E., Bhattacharya, D., *et al.* (2024) Nuclear genomes of dinoflagellates reveal evolutionarily conserved pattern of RNA editing relative to stress response. *Frontiers in Protistology* **2**

Chen, Y., González-Pech, R.A., Stephens, T.G., *et al.* (2020) Evidence That Inconsistent Gene Prediction Can Mislead Analysis of Dinoflagellate Genomes. *Journal of Phycology* **56**, 6–10

Chen, Y., Hu, Z. & Wang, H. (2019) *Muricauda amphidinii* sp. nov., a novel marine bacterium isolated from the phycosphere of dinoflagellate *Amphidinium carterae*. *International Journal of Systematic and Evolutionary Microbiology* **71**

Chen, Z., Yang, L., Li, Y., *et al.* (2014) *Cyclobacterium xiamenense* sp. nov., isolated from aggregates of *Chlorella autotrophica*, and emended description of the genus *Cyclobacterium*. *International Journal of Systematic and Evolutionary Microbiology* **64**, 887–893

Chernikova, T.N., Bargiela, R., Toshchakov, S. V., *et al.* (2020) Hydrocarbon-Degrading Bacteria *Alcanivorax* and *Marinobacter* Associated With Microalgae *Pavlova lutheri* and *Nannochloropsis oculata*. *Frontiers in Microbiology* **11**

Chin, G.J.W.L., Jani, J., Law, S.V., *et al.* (2023) Whole genome sequence data of a marine bacterium, *Marinobacter adhaerens* PBVC038, associated with toxic harmful algal bloom. *Data in Brief* **46**, 108768

Chow, M.H., Yan, K.T.H., Bennett, M.J., *et al.* (2010) Birefringence and DNA Condensation of Liquid Crystalline Chromosomes. *Eukaryotic Cell* **9**, 1577–1587

Chudnovsky, Y., Li, J.F., Rizzo, P.J., *et al.* (2002) Cloning, Expression, and Characterization of a Histone-like Protein from the Marine Dinoflagellate *Lingulodinium polyedrum* (Dinophyceae). *Journal of Phycology* **38**, 543–550

Clayton, C. (2019) Regulation of gene expression in trypanosomatids: living with polycistronic transcription. *Open Biology* **9**

Cock, P.J.A., Antao, T., Chang, J.T., *et al.* (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423

Cooper, T.A. & Ordahl, C.P. (1985) A single cardiac troponin T gene generates embryonic and adult isoforms via developmentally regulated alternate splicing. *The Journal of biological chemistry* **260**, 11140–8

Cruz-López, R. & Maske, H. (2016) The Vitamin B1 and B12 Required by the Marine Dinoflagellate *Lingulodinium polyedrum* Can be Provided by its Associated Bacterial Community in Culture. *Frontiers in Microbiology* **7**

Cuadrado, Á., De Bustos, A. & Figueroa, R.I. (2019) Chromosomal markers in the genus *Karenia*: Towards an understanding of the evolution of the chromosomes, life cycle patterns and phylogenetic relationships in dinoflagellates. *Scientific Reports* **9**, 3072

Cutignano, A., Nuzzo, G., Sardo, A., *et al.* (2017) The Missing piece in biosynthesis of amphidinols: First evidence of glycolate as a starter unit in New Polyketides from *Amphidinium carterae*. *Marine Drugs* **15**, 1–12

Dagenais-Bellefeuille, S. & Morse, D. (2013) Putting the N in dinoflagellates. *Frontiers in Microbiology* **4**

Damm, E., Kiene, R.P., Schwarz, J., *et al.* (2008) Methane cycling in Arctic shelf water and its relationship with phytoplankton biomass and DMSP. *Marine Chemistry* **109**, 45–59

Davidson, N.M., Hawkins, A.D.K. & Oshlack, A. (2017) SuperTranscripts: a data driven reference for analysis and visualisation of transcriptomes. *Genome Biology* **18**, 148

Davis, R.E. (1996) Spliced leader RNA trans-splicing in metazoa. *Parasitology Today* **12**, 33–40

De Coster, W. & Rademakers, R. (2023) NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics* **39**

de Koning, A.P.J., Gu, W., Castoe, T.A., *et al.* (2011) Repetitive Elements May Comprise Over Two-Thirds of the Human Genome. *PLoS Genetics* **7**, e1002384

Deeds, J. & Place, A. (2006) Sterol-specific membrane interactions with the toxins from *Karlodinium micrum* (Dinophyceae) — a strategy for self-protection? *African Journal of Marine Science* **28**, 421–425

Del Mondo, A., Smerilli, A., Sané, E., *et al.* (2020) Challenging microalgal vitamins for human health. *Microbial Cell Factories* **19**, 201

Delwiche, C.F. (1999) Tracing the Thread of Plastid Diversity through the Tapestry of Life. *The American Naturalist* **154**, S164–S177

DeMontigny, W. & Bachvaroff, T. (2025) The nuclear and mitochondrial genomes of *amoebophrya* sp. ex *Karlodinium veneficum*. *G3: Genes, Genomes, Genetics* **15**

Ding, W., Wang, S., Qin, P., *et al.* (2023) Anaerobic thiosulfate oxidation by the *Roseobacter* group is prevalent in marine biofilms. *Nature Communications* **14**, 2033

Dorrell, R.G. & Howe, C.J. (2015) Integration of plastids with their hosts: Lessons learned from dinoflagellates. *Proceedings of the National Academy of Sciences* **112**, 10247–10254

Dorrell, R.G. & Howe, C.J. (2012) Functional remodeling of RNA processing in replacement chloroplasts by pathways retained from their predecessors. *Proceedings of the National Academy of Sciences* **109**, 18879–18884

Dougan, K.E., Deng, Z.-L., Wöhlbrand, L., *et al.* (2023) Multi-omics analysis reveals the molecular response to heat stress in a "red tide" dinoflagellate. *Genome Biology* **24**, 265

Drongitis, D., Aniello, F., Fucci, L., *et al.* (2019) Roles of Transposable Elements in the Different Layers of Gene Expression Regulation. *International Journal of Molecular Sciences* **20**, 5755

Du, Q., Sui, Z., Chang, L., *et al.* (2016) Genome size of *Alexandrium catenella* and *Gracilariopsis lemaneiformis* estimated by flow cytometry. *Journal of Ocean University of China* **15**, 704–710

Dunn, N.A., Unni, D.R., Diesh, C., *et al.* (2019) Apollo: Democratizing genome annotation. *PLOS Computational Biology* **15**, e1006790

Durán-Riveroll, Lorena María, Juárez, O.E., Okolodkov, Y.B., *et al.* (2023) Morphological and Molecular Characterization of the Benthic Dinoflagellate *Amphidinium* from Coastal Waters of Mexico. *Phycology* **3**, 305–324

Durán-Riveroll, Lorena M., Weber, J. & Krock, B. (2023) First Identification of Amphidinols from Mexican Strains and New Analogs. *Toxins* **15**, 163

Ebenezer, T.E., Zoltner, M., Burrell, A., *et al.* (2019) Transcriptome, proteome and draft genome of *Euglena gracilis*. *BMC Biology* **17**, 11

Farhat, S., Le, P., Kayal, E., *et al.* (2021) Rapid protein evolution, organellar reductions, and invasive intronic elements in the marine aerobic parasite dinoflagellate *Amoebophrya* spp. *BMC Biology* **19**, 1

Fueyo, R., Judd, J., Feschotte, C., *et al.* (2022) Roles of transposable elements in the regulation of mammalian transcription. *Nature Reviews Molecular Cell Biology* **23**, 481–497

Gagat, P., Mackiewicz, D. & Mackiewicz, P. (2017) Peculiarities within peculiarities – dinoflagellates and their mitochondrial genomes. *Mitochondrial DNA Part B* **2**, 191–195
Galleron, C. (1984) The Fifth Base: A Natural Feature of Dinoflagellate DNA. In: Evolutionary Protistology. Springer Netherlands. pp 195–203
Galleron, C. & Durrand, A.M. (1979a) Characterization of a dinoflagellate (*Amphidinium carterae*) DNA. *Biochimie* **60**, 1235–1242
Galleron, C. & Durrand, A.M. (1979b) Cell cycle and DNA synthesis in a marine dinoflagellate *Amphidinium carterae*. *Protoplasma* **100**, 155–165
Gárate -Lizárraga, I. (2012) Proliferation of *Amphidinium carterae* (Gymnodiniales: Gymnodiniaceae) in Bahía de La Paz, Gulf of California. *CICIMAR Oceánides* **27**, 37–49
Gárate-Lizárraga, I., González-Armas, R., Verdugo-Díaz, G., *et al.* (2019) Seasonality of the dinoflagellate *Amphidinium cf. carterae* (Dinophyceae: Amphidiniales) in Bahía de la Paz, Gulf of California. *Marine Pollution Bulletin* **146**, 532–541
Gavelis, G.S., Herranz, M., Wakeman, K.C., *et al.* (2019) Dinoflagellate nucleus contains an extensive endomembrane network, the nuclear net. *Scientific Reports* **9**, 1–9
Gerath, M.W. & Chisholm, S.W. (1989) Change in Photosynthetic Capacity over the Cell Cycle in Light/Dark-Synchronized *Amphidinium carteri* Is Due Solely to the Photocycle. *Plant Physiology* **91**, 999–1005
Girardot, F., Monnier, V. & Tricoire, H. (2004) Genome wide analysis of common and specific stress responses in adult *Drosophila* melanogaster. *BMC Genomics* **5**, 74
Gómez, F. (2020) Dinoflagellates: Classification, Evolution, Physiology and Ecological Significance. Nova Science Publishers, Inc.
Gómez, F. (2012) A checklist and classification of living dinoflagellates (Dinoflagellata, Alveolata). *CICIMAR Oceánides* **27**, 65–140
Gonçalves, C. & Gonçalves, P. (2019) Multilayered horizontal operon transfers from bacteria reconstruct a thiamine salvage pathway in yeasts. *Proceedings of the National Academy of Sciences* **116**, 22219–22228
González-Pech, R.A., Stephens, T.G., Chen, Y., *et al.* (2021) Comparison of 15 dinoflagellate genomes reveals extensive sequence and structural divergence in family Symbiodiniaceae and genus *Symbiodinium*. *BMC Biology* **19**, 73
Goodman, C.A. & Hornberger, T.A. (2013) Measuring Protein Synthesis with SUnSET. *Exercise and Sport Sciences Reviews* **41**, 107–115
Gornik, S.G., Hu, I., Lassadi, I., *et al.* (2019) The biochemistry and evolution of the dinoflagellate nucleus. *Microorganisms* **7**
Green, D.H., Llewellyn, L.E., Negri, A.P., *et al.* (2004) Phylogenetic and functional diversity of the cultivable bacterial community associated with the paralytic shellfish poisoning dinoflagellate *Gymnodinium catenatum*. *FEMS Microbiology Ecology* **47**, 345–357
Guo, R., Wang, H., Suh, Y.S., *et al.* (2016) Transcriptomic profiles reveal the genome-wide responses of the harmful dinoflagellate *Cochlodinium polykrikoides* when exposed to the algicide copper sulfate. *BMC Genomics* **17**, 29
Gurevich, A., Saveliev, V., Vyahhi, N., *et al.* (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075
Hackett, J.D., Anderson, D.M., Erdner, D.L., *et al.* (2004) Dinoflagellates: a remarkable evolutionary experiment. *American Journal of Botany* **91**, 1523–1534
Hackett, J.D., Scheetz, T.E., Yoon, H.S., *et al.* (2005) Insights into a dinoflagellate genome through expressed sequence tag analysis. *BMC Genomics* **6**, 80
Hansen, P.J. (2011) The Role of Photosynthesis and Food Uptake for the Growth of Marine Mixotrophic Dinoflagellates1. *Journal of Eukaryotic Microbiology* **58**, 203–214
Haq, S., Bachvaroff, T. & Place, A. (2017) Characterization of Acetyl-CoA Carboxylases in the Basal Dinoflagellate *Amphidinium carterae*. *Marine Drugs* **15**, 149

Haq, S., Oyler, B.L., Williams, E., *et al.* (2023) Investigating A Multi-Domain Polyketide Synthase in *Amphidinium carterae*. *Marine Drugs* **21**, 425
Haro, R., Walunjkar, N., Jorapur, S., *et al.* (2024) Long-read DNA sequencing reveals the organization of the mitochondrial genome in the early-branching dinoflagellate *Oxyrrhis marina*. *Protist* **175**, 126071
Hieda, M., Sorada, A., Kinoshita, M., *et al.* (2021) Amphidinol 3 preferentially binds to cholesterol in disordered domains and disrupts membrane phase separation. *Biochemistry and Biophysics Reports* **26**, 100941
Hiller, R.G., Wrench, P.M., Gooley, A.P., *et al.* (1993) The Major Intrinsic Light-Harvesting Protein of *Amphidinium*: Characterization and Relation to Other Light-Harvesting Proteins. *Photochemistry and Photobiology* **57**, 125–131
Hjelmen, C.E. (2024) Genome size and chromosome number are critical metrics for accurate genome assembly assessment in Eukaryota. *GENETICS* **227**
Hofmann, E., Wrench, P.M., Sharples, F.P., *et al.* (1996) Structural Basis of Light Harvesting by Carotenoids: Peridinin-Chlorophyll-Protein from *Amphidinium carterae*. *Science* **272**, 1788–1791
Hold, G.L., Smith, E.A., Harry Birkbeck, T., *et al.* (2001) Comparison of paralytic shellfish toxin (PST) production by the dinoflagellates *Alexandrium lusitanicum* NEPCC 253 and *Alexandrium tamarense* NEPCC 407 in the presence and absence of bacteria. *FEMS Microbiology Ecology* **36**, 223–234
Hoppenrath, M. (2017) Dinoflagellate taxonomy — a review and proposal of a revised classification. *Marine Biodiversity* **47**, 381–403
Houdai, T., Matsuoka, S., Murata, M., *et al.* (2001) Acetate labeling patterns of dinoflagellate polyketides, amphidinols 2, 3 and 4. *Tetrahedron* **57**, 5551–5555
Hulburt, E.M. (1957) The Taxonomy of Unarmored Dinophyceae of Shallow Embayments on Cape Cod, Massachusetts. *The Biological Bulletin* **112**, 196–219
Jackson, A.P. (2007) Tandem gene arrays in *Trypanosoma brucei*: Comparative phylogenomic analysis of duplicate sequence variation. *BMC Evolutionary Biology* **7**, 54
Jackson, C.J., Gornik, S.G. & Waller, R.F. (2012) The Mitochondrial Genome and Transcriptome of the Basal Dinoflagellate *Hematodinium* sp.: Character Evolution within the Highly Derived Mitochondrial Genomes of Dinoflagellates. *Genome Biology and Evolution* **4**, 59–72
Jackson, C.J., Norman, J.E., Schnare, M.N., *et al.* (2007) Broad genomic and transcriptional analysis reveals a highly derived genome in dinoflagellate mitochondria. *BMC Biology* **5**, 41
Jaeckisch, N., Yang, I., Wohlrab, S., *et al.* (2011) Comparative Genomic and Transcriptomic Characterization of the Toxigenic Marine Dinoflagellate *Alexandrium ostenfeldii*. *PLoS ONE* **6**, e28012
Jagus, R., Bachvaroff, T.R., Joshi, B., *et al.* (2012) Diversity of eukaryotic translational initiation factor eIF4E in protists. *Comparative and Functional Genomics* **2012**
Janouskovec, J., Gavelis, G.S., Burki, F., *et al.* (2017) Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics. *Proceedings of the National Academy of Sciences of the United States of America* **114**, E171–E180
Jeong, S.E., Kim, K.H., Lhee, D., *et al.* (2019) *Oceaniradius stylonematis* gen. nov., sp. nov., isolated from a red alga, *Stylonema cornu-cervi*. *International Journal of Systematic and Evolutionary Microbiology* **69**, 1967–1973
Jia, Y., Gao, H., Tong, M., *et al.* (2019) Cell Cycle Regulation of the Mixotrophic Dinoflagellate *Dinophysis acuminata*: Growth, Photosynthetic Efficiency and Toxin Production. *Harmful Algae* **89**

Jiang, X., Liu, B., Nie, Z., *et al.* (2021) The role of m6A modification in the biological functions and diseases. *Signal Transduction and Targeted Therapy* **6**, 74
Jiang, Z., Duan, Y., Yang, X., *et al.* (2020) *Nitratireductor alexandrii* sp. nov., from phycosphere microbiota of toxic marine dinoflagellate *Alexandrium tamarense*. *International Journal of Systematic and Evolutionary Microbiology* **70**, 4390–4397
Jones, G.D., Williams, E., Haq, S., *et al.* (2024) Regulated mRNA recruitment in dinoflagellates is reflected in hyper-variable mRNA spliced leaders and novel eIF4Es. *ELife*
Jones, G.D., Williams, E.P., Place, A.R., *et al.* (2015) The alveolate translation initiation factor 4E family reveals a custom toolkit for translational control in core dinoflagellates. *BMC Evolutionary Biology* **15**, 1–12
Jones, G.D., Williams, E.P., Saddef Haq, *et al.* Beyond The Dinoflagellate Transcriptome: Variable mRNA spliced leaders and novel eIF4Es in *Amphidinium carterae*. 1–45
Judd, M. & Place, A.R. (2022) A Strategy for Gene Knockdown in Dinoflagellates. *Microorganisms* **10**, 1131
Judd, M., Wira, J., Place, A.R., *et al.* (2024) Long-Read Sequencing Unlocks New Insights into the *Amphidinium carterae* Microbiome. *Marine Drugs* **22**, 342
Jumper, J., Evans, R., Pritzel, A., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589
Jung, H.S., Jeong, S.E., Chun, B.H., *et al.* (2019) *Rhodophyticola porphyridii* gen. nov., sp. nov., isolated from a red alga, *Porphyridium marinum*. *International Journal of Systematic and Evolutionary Microbiology* **69**, 1656–1661
Kamikawa, R., Tanifuji, G., Kawachi, M., *et al.* (2015) Plastid Genome-Based Phylogeny Pinpointed the Origin of the Green-Colored Plastid in the Dinoflagellate *Lepidodinium chlorophorum*. *Genome Biology and Evolution* **7**, 1133–1140
Kato, K.H., Moriyama, A., Huitorel, P., *et al.* (1997) Isolation of the major basic nuclear protein and its localization on chromosomes of the dinoflagellate, *Oxyrrhis marina*. *Biology of the Cell* **89**, 43–52
Kim, E. & Archibald, J.M. (2010) Plastid evolution: gene transfer and the maintenance of "stolen" organelles. *BMC Biology* **8**, 73
Kim, K.-Y. & Kim, C.-H. (2007) Phylogenetic Relationships among Diverse Dinoflagellate Species Occurring in Coastal Waters off Korea Inferred from Large Subunit Ribosomal DNA Sequence Data. *ALGAE* **22**, 57–67
Kolmogorov, M., Yuan, J., Lin, Y., *et al.* (2019) Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology* **37**, 540–546
Koonin, E. V (2006) The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? *Biology Direct* **1**, 22
Kornberg, R.D. (1974) Chromatin Structure: A Repeating Unit of Histones and DNA. *Science* **184**, 868–871
Kowalska-Krochmal, B. & Dudek-Wicher, R. (2021) The Minimum Inhibitory Concentration of Antibiotics: Methods, Interpretation, Clinical Relevance. *Pathogens* **10**, 165
Kozlov, A.M., Darriba, D., Flouri, T., *et al.* (2019) RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455
Kumar, A. (2009) An Overview of Nested Genes in Eukaryotic Genomes. *Eukaryotic Cell* **8**, 1321–1329
Kwok, A.C.M., Mak, C.C.M., Wong, F.T.W., *et al.* (2007) Novel method for preparing spheroplasts from cells with an internal cellulosic cell wall. *Eukaryotic Cell* **6**, 563–567
Kwok, A.C.M., Yan, K.T.H., Wen, S., *et al.* (2024) Dinochromosome Heterotermini with Telosomal Anchorages. *International Journal of Molecular Sciences* **25**, 11312
Lai, F., Li, C., Zhang, Y., *et al.* (2025) DinoSource : A comprehensive database of dinoflagellate genomic resources. *Plant Biotechnology Journal*

LaJeunesse, T.C., Lambert, G., Andersen, R.A., *et al.* (2005) *Symbiodinium* (Pyrrhophyta) Genome Sizes (DNA Content) Are Smallest Among Dinoflagellates. *Journal of Phycology* **41**, 880–886
Lander, E.S., Linton, L.M., Birren, B., *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921
Langmead, B. & Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359
Larkin, S.L. & Adams, C.M. (2007) Harmful Algal Blooms and Coastal Business: Economic Consequences in Florida. *Society & Natural Resources* **20**, 849–859
Lauritano, C., De Luca, D., Ferrarini, A., *et al.* (2017) De novo transcriptome of the cosmopolitan dinoflagellate *Amphidinium carterae* to identify enzymes with biotechnological potential. *Scientific Reports* **7**, 1–12
Lazo-Murphy, B.M. & Peng, X. (2025) Unexpected Abundance and Gene Expression of *Polarella* from a Tropical Oxygen Deficient Zone. *Ocean-Land-Atmosphere Research* **4**
Le, Q.H., Markovic, P., Hastings, J.W., *et al.* (1997) Structure and organization of the peridinin-chlorophyll a-binding protein gene in *Gonyaulax polyedra*. *Molecular and General Genetics MGG* **255**, 595–604
Lee, C.Y.S., Putnam, A., Lu, T., *et al.* (2020) Recruitment of mRNAs to P granules by condensation with intrinsically-disordered proteins. *ELife* **9**, 1–31
Lee, T.C.-H., Chan, P.-L., Tam, N.F.-Y., *et al.* (2021) Establish axenic cultures of armored and unarmored marine dinoflagellate species using density separation, antibacterial treatments and stepwise dilution selection. *Scientific Reports* **11**, 202
Leighfield, T.A. & Van Dolah, F.M. (2001) Cell cycle regulation in a dinoflagellate, *Amphidinium operculatum*: Identification of the diel entraining cue and a possible role for cyclic AMP. *Journal of Experimental Marine Biology and Ecology* **262**, 177–197
LeKieffre, C., Spero, H.J., Fehrenbacher, J.S., *et al.* (2020) Ammonium is the preferred source of nitrogen for planktonic foraminifer and their dinoflagellate symbionts. *Proceedings of the Royal Society B: Biological Sciences* **287**, 20200620
Leonelli, S. & Ankeny, R.A. (2013) What makes a model organism? *Endeavour* **37**, 209–212
Li, C., Li, Y., Guo, J., *et al.* (2024a) Abundant mRNA m1A modification in dinoflagellates: a new layer of gene regulation. *EMBO Reports* **25**, 4655–4673
Li, C., Li, Ying, Wang, Y., *et al.* (2024b) Characterization of the enzyme for 5-hydroxymethyluridine production and its role in silencing transposable elements in dinoflagellates. *Proceedings of the National Academy of Sciences* **121**
Li, H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100
Li, H., Handsaker, B., Wysoker, A., *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079
Li, L. & Wang, C.C. (2005) Identification in the Ancient Protist *Giardia lamblia* of Two Eukaryotic Translation Initiation Factor 4E Homologues with Distinctive Functions. *Eukaryotic Cell* **4**, 948–959
Li, Y., Xu, Y. & Ma, Z. (2017) Comparative Analysis of the Exon-Intron Structure in Eukaryotic Genomes. *Yangtze Medicine* **01**, 50–64
Li, Y.I., Sanchez-Pulido, L., Haerty, W., *et al.* (2015) RBFOX and PTBP1 proteins regulate the alternative splicing of micro-exons in human brain transcripts. *Genome Research* **25**, 1–13
Liao, X., Zhu, W., Zhou, J., *et al.* (2023) Repetitive DNA sequence detection and its role in the human genome. *Communications Biology* **6**, 954
Liao, Y., Smyth, G.K. & Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930

Lidie, K.B., Ryan, J.C., Barbier, M., *et al.* (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**, 481–493
Lidie, K.B. & Van Dolah, F.M. (2007) Spliced leader RNA-mediated trans-splicing in a dinoflagellate, *Karenia brevis*. *Journal of Eukaryotic Microbiology* **54**, 427–435
Lin, Q., Shang, L., Wang, X., *et al.* (2021) Different dimethylsulphoniopropionate-producing ability of dinoflagellates could affect the structure of their associated bacterial community. *Algal Research* **57**, 102359
Lin, S. (2024) A decade of dinoflagellate genomics illuminating an enigmatic eukaryote cell. *BMC Genomics* **25**, 932
Lin, S. (2011) Genomic understanding of dinoflagellates. *Research in Microbiology* **162**, 551–569
Lin, S., Cheng, S., Song, B., *et al.* (2015) The *Symbiodinium kawagutii* genome illuminates dinoflagellate gene expression and coral symbiosis. *Science* **350**, 691–694
Lin, S., Zhang, H., Spencer, D.F., *et al.* (2002) Widespread and Extensive Editing of Mitochondrial mRNAs in Dinoflagellates. *Journal of Molecular Biology* **320**, 727–739
Litaker, R.W., Vandersea, M.W., Faust, M.A., *et al.* (2010) Global distribution of ciguatera causing dinoflagellates in the genus *Gambierdiscus*. *Toxicon* **56**, 711–730
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. *Marine Drugs* **15**
Liu, Y., Zhao, W.H., Dong, B., *et al.* (2019) Changes in intracellular and extracellular free polyamines during the growth cycle of *Prorocentrum donghaiense*. *IOP Conference Series: Earth and Environmental Science* **344**, 012060
Loehlin, D.W. & Carroll, S.B. (2016) Expression of tandem gene duplicates is often greater than twofold. *Proceedings of the National Academy of Sciences* **113**, 5988–5992
Lowe, C.D., Mello, L. V, Samatar, N., *et al.* (2011) The transcriptome of the novel dinoflagellate *Oxyrrhis marina* (Alveolata: Dinophyceae): response to salinity examined by 454 sequencing. *BMC Genomics* **12**, 519
Lysak, M.A., Berr, A., Pecinka, A., *et al.* (2006) Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. *Proceedings of the National Academy of Sciences* **103**, 5224–5229
Marasco, R., Michoud, G., Sefrji, F.O., *et al.* (2023) The identification of the new species *Nitratireductor thuwali* sp. nov. reveals the untapped diversity of hydrocarbon-degrading culturable bacteria from the arid mangrove sediments of the Red Sea. *Frontiers in Microbiology* **14**
Mariño-Ramírez, L., Jordan, I.K. & Landsman, D. (2006) Multiple independent evolutionary solutions to core histone gene regulation. *Genome Biology* **7**, R122
Marinov, G.K., Chen, X., Swaffer, M.P., *et al.* (2024) Genome-wide distribution of 5-hydroxymethyluracil and chromatin accessibility in the *Breviolum minutum* genome. *Genome Biology* **25**, 115
Marinov, G.K., Trevino, A.E., Xiang, T., *et al.* (2021) Transcription-dependent domain-scale three-dimensional genome organization in the dinoflagellate *Breviolum minutum*. *Nature Genetics* **53**, 613–617
Martínez-Mercado, M.A., Cembella, A.D., Sánchez-Castrejón, E., *et al.* (2024) Functional diversity of bacterial microbiota associated with the toxigenic benthic dinoflagellate *Prorocentrum*. *PLOS ONE* **19**, e0306108
Matsuo, E., Morita, K., Nakayama, T., *et al.* (2022) Comparative Plastid Genomics of Green-Colored Dinoflagellates Unveils Parallel Genome Compaction and RNA Editing. *Frontiers in Plant Science* **13**

McArthur, A.G., Waglechner, N., Nizam, F., *et al.* (2013) The Comprehensive Antibiotic Resistance Database. *Antimicrobial Agents and Chemotherapy* **57**, 3348–3357
McEwan, M., Humayun, R., Slamovits, C.H., *et al.* (2008) Nuclear Genome Sequence Survey of the Dinoflagellate *Heterocapsa triquetra*. *Journal of Eukaryotic Microbiology* **55**, 530–535
Medema, M.H., Blin, K., Cimermancic, P., *et al.* (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research* **39**, W339–W346
Mellert, K., Lamla, M., Scheffzek, K., *et al.* (2012) Enhancing Endosomal Escape of Transduced Proteins by Photochemical Internalisation. *PLoS ONE* **7**
Mertens, K.N., Rengefors, K., Moestrup, Ø., *et al.* (2012) A review of recent freshwater dinoflagellate cysts: taxonomy, phylogeny, ecology and palaeocology. *Phycologia* **51**, 612–619
Metegnier, G., Paulino, S., Ramond, P., *et al.* (2020) Species specific gene expression dynamics during harmful algal blooms. *Scientific Reports* **10**, 6182
Miller, T.R. & Belas, R. (2004) Dimethylsulfoniopropionate Metabolism by *Pfiesteria*-Associated *Roseobacter* spp. *Applied and Environmental Microbiology* **70**, 3383–3391
Monroe, E.A., Johnson, J.G., Wang, Z., *et al.* (2010) Characterization and Expression of Nuclear-Encoded Polyketide Synthases in the Brevetoxin-Producing Dinoflagellate *Karenia brevis*. *Journal of Phycology* **46**, 541–552
Morales-Amador, A., Molina-Miras, A., López-Rosales, L., *et al.* (2021) Isolation and structural elucidation of new amphidinol analogues from *Amphidinium carterae* cultivated in a pilot-scale photobioreactor. *Marine Drugs* **19**
Morden, C.W. & Sherwood, A.R. (2002) Continued evolutionary surprises among dinoflagellates. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 11558–11560
Morton, E.A., Hall, A.N., Kwan, E., *et al.* (2020) Challenges and Approaches to Genotyping Repetitive DNA. *G3 Genes|Genomes|Genetics* **10**, 417–430
Moszczyński, K., Mackiewicz, P. & Bodył, A. (2012) Evidence for Horizontal Gene Transfer from Bacteroidetes Bacteria to Dinoflagellate Minicircles. *Molecular Biology and Evolution* **29**, 887–892
Motone, K., Takagi, T., Aburaya, S., *et al.* (2020) A Zeaxanthin-Producing Bacterium Isolated from the Algal Phycosphere Protects Coral Endosymbionts from Environmental Stress. *MBio* **11**
Moustafa, A., Evans, A.N., Kulis, D.M., *et al.* (2010) Transcriptome Profiling of a Toxic Dinoflagellate Reveals a Gene-Rich Protist and a Potential Impact on Gene Expression Due to Bacterial Presence. *PLoS ONE* **5**, e9688
Murray, S., Flø Jørgensen, M., Daugbjerg, N., *et al.* (2004) *Amphidinium* Revisited. II. Resolving Species Boundaries in the *Amphidinium operculatum* Species Complex (Dinophyceae). *Journal of Phycology* **40**, 366–382
Murray, S.A., Garby, T., Hoppenrath, M., *et al.* (2012) Genetic Diversity, Morphological Uniformity and Polyketide Production in Dinoflagellates (*Amphidinium*, Dinoflagellata). *PLoS ONE* **7**, e38253
Nash, E.A., Barbrook, A.C., Edwards-Stuart, R.K., *et al.* (2007) Organization of the Mitochondrial Genome in the Dinoflagellate *Amphidinium carterae*. *Molecular Biology and Evolution* **24**, 1528–1536
National Centers for Coastal Ocean Science (2017) Assessing Environmental and Economic Impacts *Https://coastalscience.noaa.gov/science-areas/habs/assessing-environmental-and-economic-impacts/* (Accessed July 6th 2025).

National Oceanic and Atmospheric Administration - National Marine Fisheries Service (2024) Hitting Us Where It Hurts: The Untold Story of Harmful Algal Blooms
National Research Council (US) Committee on Developmental Toxicology. (2000). *Scientific frontiers in developmental toxicology and risk assessment*. National Academies Press. https://doi.org/10.17226/9871
Nawrocki, E.P. & Eddy, S.R. (2013) Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935
Nedashkovskaya, O.I., Vancanneyt, M., Kim, S.B., *et al.* (2010) Reclassification of *Flexibacter tractuosus* (Lewin 1969) Leadbetter 1974 and '*Microscilla sericea'* Lewin 1969 in the genus *Marivirga* gen. nov. as *Marivirga tractuosa* comb. nov. and *Marivirga sericea* nom. rev., comb. nov. *International Journal of Systematic and Evolutionary Microbiology* **60**, 1858–1863
Nguyen, J., Lara-Gutiérrez, J. & Stocker, R. (2021) Environmental fluctuations and their effects on microbial communities, populations and individuals. *FEMS Microbiology Reviews* **45**
Nguyen-Vo, T.-H., Nguyen, Q.H., Do, T.T.T., *et al.* (2019) iPseU-NCP: Identifying RNA pseudouridine sites using random forest and NCP-encoded features. *BMC Genomics* **20**, 971
Nishibori, N. & Imai, I. (2013) Polyamines control the growth of the fish-killing dinoflagellate *Karenia mikimotoi* in culture. *Harmful Algae* **29**, 10–13
Nuzzo, G., Cutignano, A., Sardo, A., *et al.* (2014) Antifungal Amphidinol 18 and Its 7-Sulfate Derivative from the Marine Dinoflagellate *Amphidinium carterae*. *Journal of Natural Products* **77**, 1524–1527
Oertlin, C., Watt, K., Ristau, J., *et al.* (2022) Anota2seq Analysis for Transcriptome-Wide Studies of mRNA Translation. pp 243–268
Okamoto, O.K. & Hastings, J.W. (2003) Novel Dinoflagellate Clock-Related Genes Identified Through Microarray Analysis. *Journal of Phycology* **39**, 519–526
Ollison, G.A., Hu, S.K., Hopper, J. V., *et al.* (2023) Physiology governing diatom vs. dinoflagellate bloom and decline in coastal Santa Monica Bay. *Frontiers in Microbiology* **14**
Olofsson, M., Robertson, E.K., Edler, L., *et al.* (2019) Nitrate and ammonium fluxes to diatoms and dinoflagellates at a single cell level in mixed field communities in the sea. *Scientific Reports* **9**, 1424
Ottesen, E.A., Young, C.R., Eppley, J.M., *et al.* (2013) Pattern and synchrony of gene expression among sympatric marine microbial populations. *Proceedings of the National Academy of Sciences* **110**
Oudet, P., Gross-Bellard, M. & Chambon, P. (1975) Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell* **4**, 281–300
Pagani, I., Chertkov, O., Lapidus, A., *et al.* (2011) Complete genome sequence of *Marivirga tractuosa* type strain (H-43T). *Standards in Genomic Sciences* **4**, 154–162
Pagliara, P. & Caroppo, C. (2012) Toxicity assessment of *Amphidinium carterae*, *Coolia* cfr. *monotis* and *Ostreopsis* cfr. *ovata* (Dinophyta) isolated from the northern Ionian Sea (Mediterranean Sea). *Toxicon* **60**, 1203–1214
Palacios, L., Arahal, D.R., Reguera, B., *et al.* (2006) *Hoeflea alexandrii* sp. nov., isolated from the toxic dinoflagellate *Alexandrium minutum* AL1V. *International Journal of Systematic and Evolutionary Microbiology* **56**, 1991–1995
Pan, D. & Zhang, L. (2008) Tandemly Arrayed Genes in Vertebrate Genomes. *Comparative and Functional Genomics* **2008**, 1–11
Parada, G.E., Munita, R., Georgakopoulos-Soares, I., *et al.* (2021) MicroExonator enables systematic discovery and quantification of microexons across mouse embryonic development. *Genome Biology* **22**, 43

Parks, D.H., Imelfort, M., Skennerton, C.T., *et al.* (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* **25**, 1043–1055

Patocka, J., Gupta, R.C., Wu, Q., *et al.* (2015) Toxic potential of palytoxin. *Journal of Huazhong University of Science and Technology [Medical Sciences]* **35**, 773–780

Peltomaa, E., Hällfors, H. & Taipale, S.J. (2019) Comparison of Diatoms and Dinoflagellates from Different Habitats as Sources of PUFAs. *Marine Drugs* **17**, 233

Pertea, M., Pertea, G.M., Antonescu, C.M., *et al.* (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* **33**, 290–295

Piovesan, A., Antonaros, F., Vitale, L., *et al.* (2019) Human protein-coding genes and gene feature statistics in 2019. *BMC Research Notes* **12**, 315

Place, A.R., Bai, X., Kim, S., *et al.* (2009) Dinoflagellate Host-Parasite Sterol Profiles Dictate Karlotoxin Sensitivity. *Journal of Phycology* **45**, 375–385

Place, A.R., Bowers, H.A., Bachvaroff, T.R., *et al.* (2012) *Karlodinium veneficum*-The little dinoflagellate with a big bite. *Harmful Algae* **14**, 179–195

Place, A.R., Ramos-Franco, J., Waters, A.L., *et al.* (2024) Sterolysin from a 1950s culture of *Karlodinium veneficum* (aka *Gymnodinium veneficum* Ballantine) forms lethal sterol dependent membrane pores. *Scientific Reports* **14**, 17998

Price, D.C. & Bhattacharya, D. (2017) Robust Dinoflagellata phylogeny inferred from public transcriptome databases. *Journal of Phycology* **53**, 725–729

Quast, C., Pruesse, E., Yilmaz, P., *et al.* (2012) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41**, D590–D596

Rae, P.M. (1973) 5-Hydroxymethyluracil in the DNA of a dinoflagellate. *Proceedings of the National Academy of Sciences of the United States of America* **70**, 1141–1145

Rambo, I.M., Dombrowski, N., Constant, L., *et al.* (2020) Metabolic relationships of uncultured bacteria associated with the microalgae *Gambierdiscus*. *Environmental Microbiology* **22**, 1764–1783

Raven, J.A. & Beardall, J. (2017) Consequences of the genotypic loss of mitochondrial Complex i in dinoflagellates and of phenotypic regulation of Complex i content in other photosynthetic organisms. *Journal of Experimental Botany* **68**, 2683–2692

Rein, K.S. & Borrone, J. (1999) Polyketides from dinoflagellates: origins, pharmacology and biosynthesis. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **124**, 117–131

Reugels, A.M., Kurek, R., Lammermann, U., *et al.* (2000) Mega-introns in the Dynein Gene DhDhc7(Y) on the Heterochromatic Y Chromosome Give Rise to the Giant Threads Loops in Primary Spermatocytes of *Drosophila hydei*. *Genetics* **154**, 759–769

Riding, J.B., Fensome, R.A., Soyer-Gobillard, M.-O., *et al.* (2022) A Review of the Dinoflagellates and Their Evolution from Fossils to Modern. *Journal of Marine Science and Engineering* **11**, 1

Robinson, M.D., McCarthy, D.J. & Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140

Rolinson, G. (1998) Forty years of beta-lactam research. *Journal of Antimicrobial Chemotherapy* **41**, 589–603

Rosati, D., Palmieri, M., Brunelli, G., *et al.* (2024) Differential gene expression analysis pipelines and bioinformatic tools for the identification of specific biomarkers: A review. *Computational and Structural Biotechnology Journal* **23**, 1154–1168

Ross-Kaschitza, D. & Altmann, M. (2020) eIF4E and Interactors from Unicellular Eukaryotes. *International Journal of Molecular Sciences* **21**, 2170
Rowan, R., Whitney, S.M., Fowler, A., *et al.* (1996) Rubisco in marine symbiotic dinoflagellates: form II enzymes in eukaryotic oxygenic phototrophs encoded by a nuclear multigene family. *The Plant Cell* **8**, 539–553
Roy, S., Jagus, R. & Morse, D. (2018) Translation and translational control in dinoflagellates. *Microorganisms* **6**, 1–25
Roy, S. & Morse, D. (2013) Transcription and Maturation of mRNA in Dinoflagellates. *Microorganisms* **1**, 71–99
Sakharkar, M.K., Chow, V.T.K. & Kangueane, P. (2004) Distributions of Exons and Introns in the Human Genome. *In Silico Biology: Journal of Biological Systems Modeling and Multi-Scale Simulation* **4**, 387–393
Sala-Rovira, M., Geraud, M.L., Caput, D., *et al.* (1991) Molecular cloning and immunolocalization of two variants of the major basic nuclear protein (HCc) from the histone-less eukaryote *Crypthecodinium cohnii* (Pyrrhophyta). *Chromosoma* **100**, 510–518
Sarai, C., Tanifuji, G., Nakayama, T., *et al.* (2020) Dinoflagellates with relic endosymbiont nuclei as models for elucidating organellogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 5364–5375
Schmuths, H. (2004) Genome Size Variation among Accessions of *Arabidopsis thaliana*. *Annals of Botany* **93**, 317–321
Scientific Frontiers in Developmental Toxicology and Risk Assessment (2000). National Academies Press.
Seoane, S., Molina-Miras, A., López-Rosales, L., *et al.* (2018) Data on the *Amphidinium carterae* Dn241EHU isolation and morphological and molecular characterization. *Data in Brief* **20**, 1–5
Seok, H.-Y., Lee, S.-Y., Sarker, S., *et al.* (2023) Genome-Wide Analysis of Stress-Responsive Genes and Alternative Splice Variants in *Arabidopsis* Roots under Osmotic Stresses. *International Journal of Molecular Sciences* **24**, 14580
Sereika, M., Kirkegaard, R.H., Karst, S.M., *et al.* (2022) Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nature Methods* **19**, 823–826
Seveno, J., Even, Y. & Le Gac, M. (2020) Strong constitutive expression divergence among strains but no evidence of differential expression associated with sexual reproduction in *Alexandrium minutum*. *Harmful Algae* **100**, 101940
Shah, S., Dougan, K.E., Chen, Y., *et al.* (2024) Massive genome reduction predates the divergence of Symbiodiniaceae dinoflagellates. *The ISME Journal* **18**
Shakya, M., Ahmed, S.A., Davenport, K.W., *et al.* (2020) Standardized phylogenetic and molecular evolutionary analysis applied to species across the microbial tree of life. *Scientific Reports* **10**, 1723
Shechter, D. & Allis, C.D. (2007) A lasting marriage: histones and DNA tie a knot that is here to stay. *Nature Reviews Genetics* **8**, S23–S23
Sheng, J., Malkiel, E., Katz, J., *et al.* (2010) A dinoflagellate exploits toxins to immobilize prey prior to ingestion. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 2082–2087
Shoguchi, E., Shinzato, C., Kawashima, T., *et al.* (2013) Draft Assembly of the *Symbiodinium minutum* Nuclear Genome Reveals Dinoflagellate Gene Structure. *Current Biology* **23**, 1399–1408

Simillion, C., Vandepoele, K., Van Montagu, M.C.E., *et al.* (2002) The hidden duplication past of *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences* **99**, 13627–13632

Song, H., Zhang, J., Liu, B., *et al.* (2022) Biological roles of RNA m5C modification and its implications in Cancer immunotherapy. *Biomarker Research* **10**, 15

Soyer-Gobillard, M.O. & Geraud, M.L. (1992) Nucleolus behaviour during the cell cycle of a primitive dinoflagellate eukaryote, *Prorocentrum micans* Ehr., seen by light microscopy and electron microscopy. *Journal of Cell Science* **102**, 475–485

Spector, D.L. (1984) *Dinoflagellate Nuclei*. in D.L. Spector (Ed), Dinoflagellates (pp. 107-147), Academic Press, Inc.

Stephens, T.G., González-Pech, R.A., Cheng, Y., *et al.* (2020) Genomes of the dinoflagellate *Polarella glacialis* encode tandemly repeated single-exon genes with adaptive functions. *BMC Biology* **18**, 1–21

Stephens, T.G., Ragan, M.A., Bhattacharya, D., *et al.* (2018) Core genes in diverse dinoflagellate lineages include a wealth of conserved dark genes with unknown functions. *Scientific Reports* **8**, 17175

Stover, N.A., Kaye, M.S. & Cavalcanti, A.R.O. (2006) Spliced leader trans-splicing. *Current Biology* **16**, 8–9

Strompl, C. (2003) *Oceanicaulis alexandrii* gen. nov., sp. nov., a novel stalked bacterium isolated from a culture of the dinoflagellate *Alexandrium tamarense* (Lebour) Balech. *International Journal of Systematic and Evolutionary Microbiology* **53**, 1901–1906

Stüken, A., Orr, R.J.S., Kellmann, R., *et al.* (2011) Discovery of Nuclear-Encoded Genes for the Neurotoxin Saxitoxin in Dinoflagellates. *PLoS ONE* **6**, e20096

Stüken, A., Riobó, P., Franco, J., *et al.* (2015) Paralytic shellfish toxin content is related to genomic sxtA4 copy number in *Alexandrium minutum* strains. *Frontiers in Microbiology* **6**

Su, J., Yang, X., Zheng, T., *et al.* (2007) An efficient method to obtain axenic cultures of *Alexandrium tamarense*—a PSP-producing dinoflagellate. *Journal of Microbiological Methods* **69**, 425–430

Summerton, J.E. (2005) Endo-Porter: A Novel Reagent For Safe, Effective Delivery Of Substances Into Cells. *Annals of the New York Academy of Sciences, 1058*(1), 62-75

Sun, H., Ding, J., Piednoël, M., *et al.* (2018) findGSE: estimating genome size variation within human and *Arabidopsis* using k-mer frequencies. *Bioinformatics* **34**, 550–557

Swarbreck, D., Wilks, C., Lamesch, P., *et al.* (2007) The *Arabidopsis* Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Research* **36**, D1009–D1014

Takagi, T., Aoyama, K., Motone, K., *et al.* (2023) Mutualistic Interactions between Dinoflagellates and Pigmented Bacteria Mitigate Environmental Stress. *Microbiology Spectrum* **11**

Takishita, K. (2004) Phylogeny of Nuclear-Encoded Plastid-Targeted GAPDH Gene Supports Separate Origins for the Peridinin- and the Fucoxanthin Derivative-Containing Plastids of Dinoflagellates. *Protist* **155**, 447–458

Tarailo-Graovac, M. & Chen, N. (2009) Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. *Current Protocols in Bioinformatics* **25**

Tarazona-Janampa, U.I., Cembella, A.D., Pelayo-Zárate, M.C., *et al.* (2020) Associated Bacteria and Their Effects on Growth and Toxigenicity of the Dinoflagellate *Prorocentrum lima* Species Complex From Epibenthic Substrates Along Mexican Coasts. *Frontiers in Marine Science* **7**

Tengs, T., Dahlberg, O.J., Shalchian-Tabrizi, K., *et al.* (2000) Phylogenetic Analyses Indicate that the 19′Hexanoyloxy-fucoxanthin-Containing Dinoflagellates Have Tertiary Plastids of Haptophyte Origin. *Molecular Biology and Evolution* **17**, 718–729

Thakur, J., Packiaraj, J. & Henikoff, S. (2021) Sequence, Chromatin and Evolution of Satellite DNA. *International Journal of Molecular Sciences* **22**, 4309

Tizabi, D., Hill, R.T. & Bachvaroff, T. (2025) Nanopore Sequencing of *Amoebophrya* Species Reveals Novel Collection of Bacteria Putatively Associated with *Karlodinium veneficum*. *Genome Biology and Evolution* **17**

Todaro, A.M., Hackeng, T.M. & Castoldi, E. (2021) Antisense-mediated down-regulation of factor v-short splicing in a liver cell line model. *Applied Sciences (Switzerland)* **11**

Town, C.D., Cheung, F., Maiti, R., *et al.* (2006) Comparative Genomics of *Brassica oleracea* and *Arabidopsis thaliana* Reveal Gene Loss, Fragmentation, and Dispersal after Polyploidy. *The Plant Cell* **18**, 1348–1359

Tse, S.P.K., Lee, F.W.F., Mak, D.Y.L., *et al.* (2020) Production of Paralytic Shellfish Toxins (PSTs) in toxic *Alexandrium catenella* is intertwined with photosynthesis and energy production. *Toxins* **12**, 1–25

Ullu, E. & Tschudi, C. (1995) Accurate Modification of the Trypanosome Spliced Leader Cap Structure in a Homologous Cell-free System. *Journal of Biological Chemistry* **270**, 20365–20369

Ullu, E. & Tschudi, C. (1991) Trans splicing in trypanosomes requires methylation of the 5' end of the spliced leader RNA. *Proceedings of the National Academy of Sciences* **88**, 10074–10078

University of Florida Sea Grant (2024) Assessment of the Short- and Long-term Socioeconomic Impacts of Florida's 2017–2019 Red Tide Event: Project Summary

Uribe, P. & Espejo, R.T. (2003) Effect of Associated Bacteria on the Growth and Toxicity of *Alexandrium catenella*. *Applied and Environmental Microbiology* **69**, 659–662

Ustianenko, D., Weyn-Vanhentenryck, S.M. & Zhang, C. (2017) Microexons: discovery, regulation, and function. *WIREs RNA* **8**

Van Dolah, F.M., Lidie, K.B., Morey, J.S., *et al.* (2007) Microarray analysis of diurnal- and circadian-regulated genes in the Florida red-tide dinoflagellate *Karenia brevis* (Dinophyceae). *Journal of Phycology* **43**, 741–752

Varadi, M., Bertoni, D., Magana, P., *et al.* (2024) AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Research* **52**, D368–D375

Verma, A., Barua, A., Ruvindy, R., *et al.* (2019) The Genetic Basis of Toxin Biosynthesis in Dinoflagellates. *Microorganisms* **7**, 222

Vico, P., Iriarte, A., Bonilla, S., *et al.* (2021) Metagenomic analysis of *Raphidiopsis raciborskii* microbiome: beyond the individual. *Biodiversity Data Journal* **9**

Viguera, E. (2001) Replication slippage involves DNA polymerase pausing and dissociation. *The EMBO Journal* **20**, 2587–2595

Wagih, O. (2017) ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics* **33**, 3645–3647

Wagner, N.L., Greco, J.A., Enriquez, M.M., *et al.* (2013) The Nature of the Intramolecular Charge Transfer State in Peridinin. *Biophysical Journal* **104**, 1314–1325

Waller, R.F. & Jackson, C.J. (2009) Dinoflagellate mitochondrial genomes: stretching the rules of molecular biology. *BioEssays* **31**, 237–245

Wang, B., Yao, M., Zhou, J., *et al.* (2018) Growth and Toxin Production of *Gambierdiscus* spp. Can Be Regulated by Quorum-Sensing Bacteria. *Toxins* **10**, 257

Wang, Y. & Coyne, K.J. (2024) Molecular Insights into the Synergistic Effects of Putrescine and Ammonium on Dinoflagellates. *International Journal of Molecular Sciences* **25**, 1306

Wang, Z., Zhang, Z., Li, C., *et al.* (2019) *Algihabitans albus* gen. nov., sp. nov., isolated from a culture of the green alga *Ulva prolifera. International Journal of Systematic and Evolutionary Microbiology* **69**, 828–832
Wick, R.R., Judd, L.M. & Holt, K.E. (2023) Assembling the perfect bacterial genome using Oxford Nanopore and Illumina sequencing. *PLOS Computational Biology* **19**, e1010905
Williams, E., Bachvaroff, T. & Place, A. (2022) Dinoflagellate Phosphopantetheinyl Transferase (PPTase) and Thiolation Domain Interactions Characterized Using a Modified Indigoidine Synthesizing Reporter. *Microorganisms* **10**, 687
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. *Evolutionary Bioinformatics* **17**
Wisecaver, J.H., Brosnahan, M.L. & Hackett, J.D. (2013) Horizontal gene transfer is a significant driver of gene innovation in dinoflagellates. *Genome Biology and Evolution* **5**, 2368–2381
Wong, J.T.Y. (2019) Architectural Organization of Dinoflagellate Liquid Crystalline Chromosomes. *Microorganisms* **7**, 27
Wu, Z., Lee, W.H., Liu, Z., *et al.* (2022) Microbiome Associated with *Gambierdiscus balechii* Cultures Under Different Toxicity Conditions. *Frontiers in Marine Science* **9**
Xiang, T., Nelson, W., Rodriguez, J., *et al.* (2015) *Symbiodinium* transcriptome and global responses of cells to immediate changes in light intensity when grown under autotrophic or mixotrophic conditions. *The Plant Journal* **82**, 67–80
Yamada, N., Bolton, J.J., Trobajo, R., *et al.* (2019) Discovery of a kleptoplastic 'dinotom' dinoflagellate and the unique nuclear dynamics of converting kleptoplastids to permanent plastids. *Scientific Reports* **9**, 10474
Yan, T.H.K., Wu, Z., Kwok, A.C.M., *et al.* (2020) Knockdown of dinoflagellate condensin CcSMC4 subunit leads to s-phase impediment and decompaction of liquid crystalline chromosomes. *Microorganisms* **8**
Yoon, H.S., Hackett, J.D. & Bhattacharya, D. (2002) A single origin of the peridinin- and fucoxanthin-containing plastids in dinoflagellates through tertiary endosymbiosis. *Proceedings of the National Academy of Sciences* **99**, 11724–11729
Yoon, J.-H., Kang, S.-J., Lee, S.-Y., *et al.* (2009) *Seohaeicola saemankumensis* gen. nov., sp. nov., isolated from a tidal flat. *International Journal of Systematic and Evolutionary Microbiology* **59**, 2675–2679
Zaheri, B., Dagenais-Bellefeuille, S., Song, B., *et al.* (2019) Assessing Transcriptional Responses to Light by the Dinoflagellate *Symbiodinium*. *Microorganisms* **7**, 261
Zaheri, B. & Morse, D. (2022) An overview of transcription in dinoflagellates. *Gene* **829**, 146505
Zhang, C. & Lin, S. (2019) Initial evidence of functional siRNA machinery in dinoflagellates. *Harmful Algae* **81**, 53–58
Zhang, D., Zhu, L., Gao, Y., *et al.* (2024) RNA editing enzymes: structure, biological functions and applications. *Cell & Bioscience* **14**, 34
Zhang, H., Bhattacharya, D. & Lin, S. (2007a) A Three-Gene Dinoflagellate Phylogeny Suggests Monophyly of Prorocentrales and a Basal Position for *Amphidinium* and *Heterocapsa*. *Journal of Molecular Evolution* **65**, 463–474
Zhang, H., Hou, Y., Miranda, L., *et al.* (2007b) Spliced leader RNA trans-splicing in dinoflagellates. *Proceedings of the National Academy of Sciences* **104**, 4618–4623
Zhou, Z., Tran, P.Q., Breister, A.M., *et al.* (2022) METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. *Microbiome* **10**, 33
Zimin, A. V., Marçais, G., Puiu, D., *et al.* (2013) The MaSuRCA genome assembler. *Bioinformatics* **29**, 2669–2677