Contact information: barrerall@appstate.edu
Cover page
Style: APA
Journal guide: Journal of Phycology

TRANSCRIPTOME ANALYSIS OF *PYROCYSTIS LUNULA TO SHED LIGHT ON BIOLUMINESCENCE*

A Thesis
by
LILA L. BARRERA

Submitted to the School of Graduate Studies
at Appalachian State University
in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE

May 2023
Department of Biology

ANALYSIS OF *PYROCYSTIS LUNULA* TO SHED LIGHT ON BIOLUMINESCENCE

A Thesis
by
LILA L. BARRERA
May 2023

APPROVED BY:

Cara Fiore, Ph.D.
Chairperson, Thesis Committee

Mark Venable, Ph.D.
Member, Thesis Committee

Andrew Bellemer, Ph.D
Member, Thesis Committee

Ava Udvadia, Ph.D
Chairperson, Department of Biology

Marie Hoepfl, Ed.D.
Interim Dean, Cratis D. Williams School of Graduate Studies

Abstract

TRANSCRIPTOME ANALYSIS OF *PYROCYSTIS LUNULA* TO SHED LIGHT ON BIOLUMINESCENCE

Lila Lynnette Barrera
B.S., Appalachian State University
M.S., Appalachian State University

Chairperson: Cara Fiore, Ph.D

Dinoflagellates are planktonic eukaryotes abundant in many aquatic habitats, especially within marine environments. Often seen at night along the shore, some species of dinoflagellates produce light when disturbed in a process known as bioluminescence. Though fundamental work in this field has shown the importance of the circadian rhythm in the regulation of bioluminescence, there are few examples that explore gene regulation in response to this process. Gene expression studies on these species could improve our understanding of their physiology and evolutionary relationships. Here, we have built upon previously published work by leveraging circadian regulation and RNA sequencing analysis to further investigate the biochemical pathway of bioluminescence within the marine dinoflagellate, *Pyrocystis lunula*. The aim is to compare differential gene expressions of proteins involved in the signaling pathway of bioluminescence based on their circadian rhythm. We are particularly interested in the components utilized in the initiation of the pathway, particularly, mechanoreceptors. Mechanoreceptors are poorly characterized compared

to chemoreceptors and dinoflagellate bioluminescence provides an avenue for exploring potential novel mechanoreceptors. To achieve this differential gene expression, cultures were grown separately in inverse light conditions. Harvesting occurred during the late exponential growth phase while one set of cultures was in the light phase and the other in the dark phase. RNA sequencing and subsequent transcriptome analysis have yielded genes of interest that are up-or downregulated in the cultures harvested in the dark compared to those harvested in the light. Notable systems that displayed upregulated genes in the dark phase compared to the light phase included cytoskeletal regulation, circadian rhythm, and fluid shear stress annotated using KEGG. This work has the potential to advance our understanding of the cellular mechanisms involved in bioluminescence and specific insight into eukaryotic microbial mechanosensors.

Acknowledgments

I would like to express my deepest appreciation to Dr. Cara Fiore without whom the person and scientist that I have become would not exist. To Mark Venable, thank you for welcoming me and my dinoflagellates into your lab, for your endless wisdom, and your dry humor. To Andrew Bellemer, thank you for your sound guidance and silent yet loud confidence in my success. To Reis Gadsden, thank you for your hours of hard work, patience, and mentorship in the field of computer science. Many thanks to the Office of Student Research at Appalachian State University for its financial support throughout the years. To Erin Salzlein, my friend and editor, I thank you for your never-ending fervor to consume and destroy my writing in the best way possible. To my friends and family, there aren't enough words to string together to show my appreciation for all that you have done to support me in this journey.

Table of Contents

List of Tables and Figures

1. Introduction

1.1 Introduction to Dinoflagellates

Dinoflagellates are single-celled, aquatic, eukaryotic organisms that are considered primary producers. These protists are diverse, residing in both marine and freshwater environments, with varying trophic abilities (i.e., photosynthesis and/or heterotrophy), genome size, and morphologies (Gomez 2020). Some dinoflagellates produce toxins and can form blooms under conditions of high nutrient input. These are known as harmful algal blooms (HABs), such as the red tide caused by dinoflagellate species such as *Karenia brevis* or *Lingolodinium polyedra* (Anderson et al. 2021).

Dinoflagellates are also responsible for the striking blue glow sometimes visible in crashing ocean waves. This phenomenon is called bioluminescence hypothesized to be a defense mechanism for some dinoflagellates, like that of toxin production. Bioluminescence has been identified in 17 dinoflagellate genera such as *Gonyaulaux*, *Pyrocycstis*, *Noticula* and is surprisingly common throughout the kingdoms of life (Valiadi and Iglesias-Rodriguez 2013, Wilson and Hastings 2013). These species emit light in response to mechanical stress upon their cellular membranes, such as a crashing wave or a brush with a predator. This response is thought to be a communicative action that aids anti-predation in an altruistic fashion, by way of illuminating the predator to attract a secondary predator capable of eating the first (Widder 2010, Valiadi and Iglesias-Rodriguez 2013). The general chemical pathway of bioluminescence is well-documented in fireflies, consisting of the oxidation of the luciferin

protein catalyzed by the enzyme luciferase (de Wet et al. 1985). In dinoflagellates specifically, bioluminescence is circadian regulated, occurring only at night and often accompanied by luciferin binding protein (LBP) that interacts with luciferin. However, the mechanoreceptor in the dinoflagellate cell membrane that initiates the bioluminescence pathway is not known and if characterized, it would provide insight into a potentially novel type of mechanoreceptors and/or insight into the evolution of mechanoreceptor protein families.

1.2 Circadian Rhythm and Bioluminescence

Bioluminescence occurs primarily in the dark part of the light/dark cycle. Many other cellular processes are also impacted by sunlight, such as photosynthesis or temperature regulation (Serin and Acar Tek 2019). These cellular processes follow the diel cycle which is approximately a 24-hr cycle. The ability of an organism to maintain this pattern is referred to as a "circadian rhythm" or an "internal clock." Some studies show that the rhythms are sensitive to environmental stimuli such as pulses of light (McClung 2006). In fact, an organism's circadian rhythm can shift when consistently exposed to blue light or short-wave light (Bryk et al. 2022, Hittle and Wong 2022). In plants, environmental time cues, such as light/dark cycles and temperature, drive the endogenous timing and these are easily manipulated within a laboratory setting (McClung 2006)

Due to this circadian regulation, cellular components involved in bioluminescence are also circadian regulated. This looks different depending on the species. For instance, *L. polyedra* synthesizes and destroys the necessary components for bioluminescence daily, resulting in a decrease in the concentration of those cellular components during the light phase to less than

10 percent of the concentration measured during the dark phase. In contrast, *P. lunula* maneuvers the cell components to the center of the cell during the light phase and to the cell's periphery during the dark phase (Hastings 2013, Wilson and Hastings 2013).

1.3 Dinoflagellate Bioluminescence

While the bioluminescence pathway in dinoflagellates is not completely characterized there are several important steps within the cascade that are documented. *P. lunula* structure can be separated into two major parts: the central region and the periphery. The central region contains the nucleus and plastids during the evening and scintillons during the day, while the periphery contains vacuoles (Seo and Fritz 2000). Genera *Noctiluca* and *Pyrocystis* house a central vacuole that is hypothesized to interact with scintillons during the bioluminescent cascade (Johnson et al. 1985). When a mechanical force is applied to the dinoflagellate's cellular membrane, it triggers a transmembrane receptor, proposed to be a G-coupled protein receptor (GPCR) (Fig. 1). Intracellular Ca^{2+} concentration increases, allowing for electric signaling to occur across the vacuolar membrane. As a scintillon (containing components necessary for bioluminescence) projects into the vacuole, the generated action potential travels across the scintillon membrane. As a result, voltage-gated proton channels are opened which facilitates the acidification of the scintillon, creating an ideal environment for bioluminescence to occur (Valiadi and Iglesias-Rodriguez 2013).

Dinoflagellate bioluminescent machinery resides in the scintillons, which migrate to the periphery of cell and back towards the nucleus (Fig. 2). Proteins luciferin, luciferase, and luciferin binding protein (LBP) are the most common participants in dinoflagellate

bioluminescence (Seo and Fritz 2000). LBP is often utilized to prevent unwanted oxidation of luciferin. While the genetic information to create LBP is available to *P. lunula*, there is little evidence that LBP is necessary for luminescence in this species (Valiadi and Iglesias-Rodriguez 2013, Fajardo et al. 2019). In most species, bioluminescence occurs when the substrate luciferin, in the presence of oxygen, is oxidized by the enzyme luciferase to make oxyluciferin. As electrons in oxyluciferin fall from a higher quantum energy state to a lower state, light emission occurs (Rodriguez et al. 2017, Fajardo et al. 2020, Wang and Liu 2020).

1.4 Introduction to mechanoreceptors

Generally, mechanotransduction classifies processes by which an organism can process and respond to mechanical stimuli (Paluch et al. 2015). Mechanical stimuli include stretch/strain, compression, or shear stress across a cell membrane. Once applied to the cell membrane, these types of stimuli activate several types of cellular components and pathways, including G-protein coupled receptors (GPCR), ion channels, and other proteins involved in both intracellular and extracellular communication (Goldmann 2014). It is not yet known what type of mechanoreceptor receives the initial stimuli in dinoflagellates, but some well-studied examples and potential candidates include degenerin/epithelial sodium channels (DEG/ENaC), transient receptor potential (TRP) channels, and GPCRs.

DEG/ENaC are responsible for a cell's response to several kinds of stimuli including acidity, sodium (Na^+) imbalance, mechanical stimuli, and nociceptive pain (Mobasheri et al. 2005). More specifically, ENaC has been shown to be activated by laminar shear stress (LSS)

in select cells of frogs and rabbits (Bhalla and Hallows 2008). This is of interest in the dinoflagellate system due to the shear stress initiation of bioluminescence. However, a survey of dinoflagellate transcriptomes found only one out of 27 transcriptomes contained DEG/ENaC homologs (Pozdnyakov et al. 2021). The limited presence in dinoflagellate transcriptomes suggests low prevalence of these channels in dinoflagellates. Within the same transcriptome study, TRP channels were identified to be involved in mechanosensing in dinoflagellates (Pozdnyakov et al. 2021). Thus, the bioluminescent cascade in dinoflagellates may behave like that of TRP-mediated pathways in other organisms, such as TRPV in the roundworm *C. elegans*, where TRP is activated downstream of GPCRs (Kahn-Kirby et al. 2004).

There is evidence in model organisms, such as in *Drosophila* phototransduction, that TRP channels and GPCRs work in conjunction to perform mechanotransduction (Montell 2012). This may be similar to how bioluminescence in dinoflagellates initializes and if so, there may be an increase in transcripts for TRP and/or GPCR proteins in dinoflagellates that are in the dark phase of the diel cycle. Generally, when an agonist or ligand activates a GPCR, it causes a conformational change of the protein. The conformational change allows interaction between GPCR and an associated G-protein, thereby activating the G-protein. The G-protein activate enzymes and ion channels, which allow ions – in the case of bioluminescence, Ca^{2+}ion - to act as secondary messengers (Vauquelin and von Mentzer 2007, Rosenbaum et al. 2009, Shalaeva et al. 2015).

Once initialized, the activation of the GPCR would allow an action potential across the membranes of vacuoles and scintillons (Von Dassow and Latz 2002, Chen et al. 2007). The

next critical step in the pathway is cytosolic Ca^{2+} concentration increase and the release of intracellular Ca^{2+} stores. The activation of ion channels via GPCRs may be a key step to achieve the necessary high Ca^{2+} concentrations. G-proteins are also known to recruit Ca^{2+} as secondary messengers in various transduction pathways and may also perform these important roles in bioluminescence by initializing downstream processes yet to be identified (Ma et al. 2017, Pfeil et al. 2020). However, mechanoreceptors involved in bioluminescence are some of the fastest mechano-sensitive systems known with a response delay between 15-20 ms (Valiadi and Iglesias-Rodriguez 2013, Fajardo et al. 2020). Although fast, GPCRs in *Drosophila* have a latency time of 20ms, which is longer than the complete process of converting physical stimuli to light emission in dinoflagellates (Hardie and Raghu 2001, Valiadi and Iglesias-Rodriguez 2013). Thus, even though receptors between *Drosophila* and *Pyrocystis* may differ greatly, the time discrepancy suggests there may be different receptor types involved in bioluminescence or that there is a combination of proteins involved in initiating bioluminescence. Both GPCRs and TRP channels have been shown to contribute to the bioluminescence pathway in other dinoflagellates, creating a promising connection between these components (Chen et al. 2007, Lindström et al. 2017). These channels could contribute to a quicker dispersion of secondary messengers, allowing for that 20 ms reaction time (Valiadi and Iglesias-Rodriguez 2013). As shown in the conversion of phosphatidylinositol-4, 5-bisphosphate (PIP_2) to inositol-2,4,5-triphosphate (IP_3), a family of TRP channels are activated downstream of a GPCR (Inanobe and Kurachi 2014). Additional associations to be considered in characterizing the bioluminescence pathway are how scintillons are moved within the cell and how they interact with vacuole membranes.

In an effort to investigate potential proteins involved in this bioluminescence cascade, this study utilizes RNA sequencing on *P.lunula* cultures that were grown on an opposing light/dark cycle. Sequencing the mRNA transcripts provides a snapshot of what is occurring in the cell at the time of RNA extraction. Given that the mechanosensitive protein responsible for bioluminescence may be a GPCR, cofactors common to GPCR function or the GPCR itself may be upregulated in the dark phase compared to the light. Comparing dinoflagellate cultures between their light and dark phases, where bioluminescent responses may differ, would allow for the identification of key pieces of the transduction cascade if there is a change in transcript abundance. Often these cofactors are used broadly but should still be considered a target of interest during analysis of the RNAseq data. Specifically, by comparing similar functioning GPCRs to the *P. lunula*-derived sequences I expect to elucidate basic functioning and interactions of these target proteins.

Due to dinoflagellates' large genome size (1 - 270 giga base pairs (Gb), Hong et al. 2016) there are few published genomes and no data yet on circadian gene expression (Fajardo et al. 2019). The large genome size hinders genome sequencing, but RNA sequencing provides a more accessible method of insight because the transcriptome is smaller than the genome. My objective in this study was to leverage the circadian nature of bioluminescence by comparing the transcriptome of *P. lunula* cultures under light and dark conditions. I hypothesize that genes that encode for proteins involved in bioluminescence will be more highly expressed under dark phase conditions. Secondly, I anticipate being able to identify novel involved in bioluminescence by comparing the transcriptome of *P. lunula* between our light and dark conditions. These genes will then be used in future experiments for further validation. I also

predict that there is a novel protein within the cell membrane that initializes the bioluminescent cascade. Either the receptor is indeed a GPCR – faster than known mechanosensitive receptors or there are additional cellular components present that speed up the response-time.

2. Laboratory Methods

2.1 General set-up

Cultures of *P. lunula* (CCMP731) were purchased from Bigelow National Center for Marine Algae and Microbiota (NCMA) and grown in F/2 seawater media made in 0.2 μm filtered seawater from the South Carolina coast. An F/2 media kit from Bigelow NCMA was used to create the media following the manufacturer's protocol. In order to create independent light environments, two adjacent chambers were created using thick black plastic that covered the experimental area. The chambers were supported by metal shelves with sections of the black plastic separating the two chambers (Figure 3). This setup blocked most external light and did not allow light from either chamber to leak into the other. The chamber unit was set up in a culture room that remained dark aside from short periods of light used to maintain the cultures. Four LED lights were positioned and fastened to the top of each chamber, facing the flasks. The light level was tested using a photometer with an average of 50 μEinsteins, which is within the range suggested by Bigelow (40 - 120 μEinsteins). Lights were connected to adjustable timers and set to a 13-hour light period and an 11-hour dark period, opposite each other. Replicate cultures (n = 5 per treatment) were grown in these chambers at room temperature for four weeks prior to beginning the experiment and an additional four weeks

during the experiment. The cultures were initially set up using a final cell concentration of 4.85E3 cells/mL of *P. lunula* cells in the stationary phase. Cultures were grown statically for the duration of the experiment.

2.2 RNA extraction

Cultures from each treatment were harvested in the middle of either light or dark phase by a series of gentle centrifugations at 2000 x g for 30 seconds. The light treatment cultures were harvested first, followed by the dark treatment cultures with minimal light during the harvest. Cell pellets were resuspended in a phenol lysis buffer and pipetted directly into tubes with glass beads for homogenization. The Qiagen RNeasy Plus Universal Mini kit was used post-homogenization following the manufacturer protocol. Qubit fluorometric concentrations were measured for each sample. The lowest-scoring replicate from each condition was excluded from the RNA sequencing step. In total, 8 replicates of total RNA were sent to Genewiz (Azenta) for high throughput sequencing. Enrichment of mRNA was performed by polyA selection followed by cDNA synthesis for library generation. The library was then sequenced using Illumina HiSeq 2000 with reads of paired end sequencing of 150 bp in length.

2.3 Quality control

An online computing platform, Galaxy, was utilized initially for quality control (QC) using the tools FastQC and Trimmomatic (Afgan et al. 2022). Parameters for the Trimmomatic tool were as follows: 1) standard, paired-ended adaptor sequences for HiSeq were selected, 2)

the maximum mismatch count which will allow a full match to be performed that was allowed was 2 mismatch counts, 3) an accuracy score of 30 was required for a match between two adapted ligated reads for a paired-ended palindrome read alignment, 4) an accuracy score of 10 was required for the match between an adapter and a read, and 5) the minimum length of an adapter that needs to be detected was set to a length of 8, to keep both forward and reverse reads. Additionally, for quality trimming, bases were removed at the end of a read if they were below a threshold quality of 3 for both the trailing and leading trims. Within the sliding window trimming, the number of bases that average across was 4 and the average quality score required for the read to be included in analysis was set to 20. Reads that scored an average quality score lower than 25 and reads that were below the minimum length of 50 base pairs were removed. The raw RNA sequencing data of *P. lunula* will be deposited in the NCBI database.

2.4 Transcriptome assembly

The filtered and trimmed RNA reads obtained from Galaxy were then further analyzed using a high-performance computing server (HPC) at the University of North Carolina (UNC) at Chapel Hill. The cleaned files were used to assemble transcripts, using the de-novo assembly tool, Trinity (Grabherr et al. 2011). Trinity is a de novo assembly tool that reconstructs probable transcripts from raw RNAseq data. Trinity uses KMER, a substring of nucleotides in an RNA sequence, and de Bruijn graphs, graphs that represent overlaps between sequences, to reconstruct transcripts. Default settings for paired end mode were utilized for analysis. A transcriptome assembly from the raw sequence reads was performed to identify transcripts that showed significant differences in abundance. The raw sequencing data consists of 8 samples,

each with forward and reverse reads equaling 16 output files. Assembly was completed using the UNC's HPC server, Longleaf, using the Unix/Linux language. Our current methodology largely followed the procedure that Logan Whitehouse of UNC outlines in their Diatom Metatranscriptomes repository (Whitehouse 2019). Following the assembly, BUSCO assessment of assembly quality was performed (Simão et al. 2015). To quantify the abundances for downstream analysis transcript counts across all samples were quantified using Salmon, a transcript-level quantification tool (Patro et al. 2017). Salmon performs a realignment by aligning each sample's sequence to the de novo assembly and quantifying the sequence hits to the assembly.

With the de novo assembly, similar contigs were clustered together. This was performed to remove redundancy and improve alignment by removing the chance of multi-mapping due to poorly reconstructed transcripts. Clustering of transcripts was completed using CD-HIT-EST from the CD-HIT suite (Huang et al. 2010). Contigs were clustered at 98% similarity and a based standard word size of 10. The clustered assembly was then used for taxonomic and functional annotation using the PhyloDB and Keggannot databases. A modified clustered assembly, containing only eukaryotic hits, was used for Interproscan annotations.

2.5 Taxonomic and functional annotation

Two genomic databases, Kyoto Encyclopedia of Genes and Genomes (KEGG) and PhyloDB, were used to annotate our assembly for both functional and taxonomic annotation, respectively (Kanehisa 2000, Whitehouse 2019). Using NCBI BLAST, we mapped our

sequences against the database using the alignment software Diamond (Buchfink et al. 2021) KEGG files provided on the UNC Longleaf server were used to build our database. We used a modified version of the program Keggannot (Chris Berthiaume 2013). The modifications skip over unresolved pathways and allow the program to run to completion. The output of the Keggannot step was an annotated fasta file as well as a tsv (tab separated value) file, containing the annotation information for each of the reconstructed contigs. Taxonomic annotation used PhyloDB, a marine life database used for metagenomics. Similar to the KEGG annotation steps, Diamond was used to construct a database. We then used the PhyloDB Annotation program to annotate our transcripts (Whitehouse 2019). Additionally, a modified version of the clustered assembly, containing eukaryotic hits only, was run through OmicsBox software, which facilitates transcript annotation against multiple databases. The software was utilized to perform NCBI BLAST nr, GO annotation, and an Interpro scan was performed on 20,569 transcripts. Output from this data was used to create overview graphs for taxonomic and functional annotation. KEGG automatic annotation server (KAAS) is associated with the KEGG database and was used to obtain is used for quick annotation and provides hierarchical categorization of KEGG pathways for transcript data (e.g., metabolism: carbohydrate metabolism: glycolysis).

2.6 Differential gene expression analysis

Salmon quantification results were utilized to perform differential gene expression analysis (DEseq), following a Bioconductor vignette (Love et al. 2014). Bioconductor projects

include free bioinformatics packages, specific to Rstudio, for use in analyzing biological assays. DEseq results were used to create the phegg heatmap (a naming combination of phylodb and Keggannot), pathview maps, PCA plot, and MA plot (M for minus in the log scale and A for average in the log scale) (Fig. 8)

Using Rstudio, metabolic and suspected bioluminescent-related pathways were created using abundance log fold change (LFC) from our DEseq data. Scripts used in Rstudio are provided in our github repository: Pyrocystis_lunula_Transcriptome_Analysis. Select proteins highlighted within these pathways were pinpointed using the phegg heatmap, their associated transcript names and KOs were obtained and used to BLAST within the NCBI nonredundant (nr) database to corroborate annotation.

2.7 Identifying contigs of interest

Contigs of interest were identified in three ways using different plots. 1) if the contig was visually different between treatments in the heatmap of log fold change of transcripts, where at least 3 out of the 4 replicates in one treatment were up- or downregulated compared to the other, 2) visualization of significant contigs based on log fold change using a volcano plot (represents log fold change and magnitude of p-values) and an MA plot (represents log fold change and relative abundance), and 3) mapping of contigs to KEGG pathways relevant to bioluminescence. Pathways of interest included those involved in signal transduction, cytoskeleton proteins, transport proteins, and photosynthesis proteins. For these specific contigs of interest, an R-based search function was used to obtain the corresponding mRNA

sequence. These contigs were then analyzed individually using NCBI BLAST nr nucleotide database to identify the likely annotation for that transcript and using the NCBI conserved domain tool to identify protein domains. The top hits of proteins were recorded and compared against the functional annotation (KEGG) previously performed.

3. Results

3.1 *Pyrocystis lunula* transcriptome sequencing and assembly

A total of 372,218,882 raw reads were obtained using paired-end sequencing on an Illumina HiSeq. Total raw reads for both light and dark treatments amounted to 180,808,107 and 191,183,426, respectively. A total of 311,263,125 high-quality reads were used in the de novo assembly, yielding 1,550,130,829 unique KMERs of size 25 (KMER = unique nucleotides for a specified length) and 421,494 total contiguous sequences (contigs) produced. Trinity also differentiates isoforms, which are generated from alternative splicing of mRNA. Based on the longest isoform per gene, the Trinity assembly produced an average contig length of 703 base pairs with an N50 (the sequence length of the shortest contig at 50% of the total assembly length) of 1291 (Table 1). Including all of the contigs produced, the Trinity assembly produced an average contig length of 822 base pairs with an N50 score of 1466. Transcriptome completeness, as measured with the program BUSCO (Simão et al. 2015) resulted in the following scores Complete: 85.0% [Complete and single copy: 52.5%, complete and duplicated : 32.5%], Fragmented: 6.7%, Missing: 8.3%, Total BUSCO groups searched: 255.

3.2 Taxonomic and functional overview of the *P. lunula* Transcriptome

The clustered assembly of contigs, a separate assembly of the original transcripts grouped based on similarity, was used for taxonomic and functional annotations. Annotation yielded a total of 398,133 transcripts for both taxonomic and functional hits against the databases PhyloDB and KEGG, respectively. BLAST hits on the clustered assembly resulted in 139,212 hits with E-values lower than 1E-10 out of a total of 152,820 (91.1%) hits. A representative sample of our sequencing data used for initial taxonomic annotation against the NCBI nr database and mapped 90% of our reads to Eukaryotic hits and 72% of the sample to the class Dinophyceae (Fig. 4). The remaining percentages mapped to bacterial or archaeal hits, which were filtered and excluded post annotation. A second, more targeted annotation using the entire modified clustered assembly against NCBI nr revealed that within the class Dinophyceae most contigs mapped to the genera *Polarella* and *Symbiodinium*. Out of 157,241 species hits, 42,902 (~ 27.2%) mapped to *Polarella*, 93775 (~59.6%) to *Symbiodinium*, and 255 (~0.16%) to *Pyrocystis lunula* (Fig. 5).

Annotation of the clustered assembly by the program Interpro, revealed that the top 20 protein domains were represented by multifunctional proteins involved in cellular metabolism and those necessary for genetic information processing (Fig. 6). The most abundant protein domains included protein kinase, acyl transferase, and cyclic nucleotide-binding domain. Similarly, annotation of the clustered assembly against KEGG Automatic Annotation Server (KAAS) revealed that most transcripts were classified within protein kinase families.

3.3 Differences in gene expression between light and dark treatments

Principal component analysis (PCA) supported two distinct populations defined by the light and dark treatments (Fig. 7). Separation of light and dark treatments were statistically supported with a permanova test (PERMANOVA, $F_{6,7} = 1.69$, $p = 0.031$). Dispersion between light and dark treatments was not significant ($F = 1.69$, $p = 0.165$). Figure 8, an MA plot, shows the log fold change (LFC) between the two conditions over the mean of normalized counts (# of hits per transcripts for each sample). Blue triangles indicate transcripts with p-values less than 0.05. The MA plot provides a broad overview of transcripts that are considered to be up- or downregulated based on their LFC.

Transcripts that mapped to the central metabolic pathways supported a decrease or maintenance state of photosynthesis and the central carbon pathways of the tricarboxylic acid (TCA) cycle in the dark treatment relative to the light treatment (Figures 9 and 10). The photosynthetic pathway indicated down-regulation in the dark of most genes with a few exceptions such as transcripts for the D1 protein (*psbA* gene), cp47, and *psbI*. Similarly, within the TCA cycle, only 3 unique genes were upregulated including fumarate hydratase, malate dehydrogenase, and citrate synthase. Transcripts that mapped to the TCA pathview map were mostly downregulated with 19 genes in some form of downregulation and 3 upregulated. Those that were upregulated were citrate synthase, fumarate hydratase (FUM1), and malate dehydrogenase (MDH1).

Contigs were also mapped to metabolic pathways that may be relevant to bioluminescence and mechanosensing, including cyclic adenosine monophosphate (cAMP), circadian rhythm, fluid shear stress, and actin cytoskeletal pathways. Within the cAMP pathway, upregulated genes in the dark treatment included an G-protein-coupled receptor (GPCR), N-methyl-D-aspartate receptor (NMDAR), adenylate cyclase (AC), β subunit of cyclic nucleotide gated channel (CNGC), and an α-C subunit of calcium voltage-gated channel (VDCC) were more prevalent; the latter three being involved in the movement of Ca^{2+}(Fig. 11). NMDAR is a glutamate receptor, a nonspecific cation channel that allows the flow of Na^{+} and K^{+} and sometimes Ca^{2+} (Purves et al. 2001). Of the transcripts that matched to G-protein α subunits, a further BLAST analysis revealed a conserved domain matched to Gα super family, corroborating the GPCR connection. Additionally, one transcript that mapped the membrane protein VDCC, and a conserved domain search with this transcript indicated the presence of an extracellular sensory domain, CHASE2.

Pathview results for circadian rhythm reveals a majority of genes are upregulated if impacted at all. The only protein that is noticeably downregulated is the clock circadian regulator (CLOCK), while its associated gene BMAL1 is upregulated (Fig. 12). Fluid shear stress pathway (Fig. 13) revealed that most genes were downregulated, 30 genes downregulated and 19 genes upregulated. Genes of interest that were upregulated include b-catenin, ERK1, IL-IR, and PI3K. The fluid shear stress pathway was mostly downregulated in the dark treatment, but with several notable upregulated genes including b-catenin, vascular endothelial cadherin (VE-Cad), interleukin-1 receptor (IL-1R), extracellular signal-regulated kinases (ERK), and phosphoinositide 3-kinase (PIK3). Lastly, the regulation of the actin

cytoskeleton (Fig. 14) pathway has a relatively even mixture of up- and downregulated genes with 19 upregulated and 16 downregulated genes in the dark phase compared to the light. Some upregulated genes of interest include subunit 5 of the actin-related protein 2/3 complex (Arp2/3), profilin 3 (PFN), and moesin (ERM).

Expression patterns of differentially expressed genes (DEGs) were visually distinct between treatments (Fig. 14). A subset of transcripts from the top 50 DEGs (Fig. 15, Table 2) were manually annotated using the BLAST search tool and corresponded to the following proteins of dinoflagellate *Symbiodinium minutum*: a transport protein, ATP synthase subunit, and an equal homology to both a kinesin-like protein and chorismate mutase.

4. Discussion

4.1 Taxonomic and functional overview of the *P. lunula* transcriptomes

The majority (72%) of taxonomic hits of the RNAseq-derived reads were identified to be of dinoflagellate origin. The cultures were not axenic and thus there were likely some transcripts from bacteria as well as some transcripts from the dinoflagellate chloroplast that would be assigned to cyanobacteria. Within dinoflagellates, relatively few reads mapped to *P. lunula* likely due to the few transcriptomes of *P. lunula* in NCBI. In comparison, the well-studied groups like *Symbiodinium* sp., the coral symbiont and *Polarella*, a free living dinoflagellate found primarily in polar regions have higher numbers of available genomes and transcriptomes. At the time of this writing, NCBI nucleotide database houses 385 entries for

the genus *Pyrocystis* and 218,398 entries for *Polarella*, and 1,259,775 entries for *Symbiodinium*. Thus, the transcript data generated here are indeed mostly dinoflagellate in origin and can be interpreted as transcripts derived from *P. lunula*.

An overview of protein domains and families revealed that the majority of highlighted proteins are involved in regular maintenance of a cell. Many of the protein families have important roles in cAMP and cGMP signaling pathways, particularly the top three most abundantly present families: protein kinase A (PKA), mitogen-activated protein kinase (MAPK), and protein kinase G (PKG). Protein kinases overall are enzymes that catalyze reactions by transferring phosphate between substrates - ATP to ADP or GDP to GTP (Cheng et al. 2011). These protein kinases are further classified according to the amino acid residue that they phosphorylate (Hanks and Quinn 1991, Martin et al. 2010). Protein kinase functions vary greatly and include major events such as cell cycling, metabolism (Cheng et al. 2011). Thus, the majority of transcripts captured here are likely part of regular cell metabolism and maintenance in the dinoflagellate cultures. Clear support for an upregulated transcript that may function as part of the mechanoreceptor involved in bioluminescence was not obtained. However, transcripts of interest in the diel metabolism of dinoflagellates and potential transcripts involved in the bioluminescence pathway were identified and are discussed below.

4.2 Changes in metabolism over the diel cycle

Separation of the two treatments was supported by ordination and statistical analysis, indicating that these populations can be used to infer differences in up- or downregulated genes

in the dark treatment discussed from hereon. The dinoflagellate cultures in the dark responded metabolically as expected, with a decrease in expression of most photosynthetic genes and maintenance of central carbon metabolic pathways such as the TCA cycle. However, there were a few genes, *D1/psbA* and *psb1*, in the photosynthesis pathway that were upregulated in the dark. These two genes are not actually regulated by circadian rhythm in the model dinoflagellate *Gonyaulax*, and examination of the transcript abundance showed no change (Wang et al. 2005). Additionally, *psbA* codes for a highly sensitive protein that degrades rapidly, leading to a high turnover rate (Wang et al. 2005, Mulo et al. 2009). The upregulation of the *psbI* gene during the dark phase may be explained by the fact that *psbI* in the cyanobacteria *Synechocystis* sp. acts as an early assembly partner for *psbA* and therefore may need to be replenished just as rapidly (Mulo et al. 2009).

Transcripts selected from the differential gene expression analysis included a subunit of a transport protein called Sec61, a subunit of ATP synthase, and a kinesin-like protein, KIF13B. Sec61 is a ubiquitous membrane protein complex that transports polypeptides but they also allow Ca^{2+} leakage from the endoplasmic reticulum to the cytosol (Lang et al. n.d., Chang and Fu 2023). Due to the transport of polypeptides, an ortholog of the sec61 channel gene is unlikely to be part of the bioluminescent cascade in a typical manner. However, the release of Ca^{2+} from organelles is a method in which dinoflagellates increase intracellular Ca^{2+} for during the bioluminescent cascade (Valiadi and Iglesias-Rodriguez 2013). Similarly, the ATP synthase is likely part of central metabolism and may be upregulated in the dark due to heterotrophic activity. A kinesin-like protein, in contrast, is a known motor protein in humans (Manning et al. 2007). The gene KIF9, encodes for Klp1, which facilitates ciliary motility in

Chlamydomonas (Konjikusic et al. 2023). The upregulation of KIF13b suggests the presence of a homologous gene that acts in a similar manner within dinoflagellates. The movement of scintillons between the cell center and the periphery is one way in which *P. lunula* regulates bioluminescence (Seo and Fritz 2000), so rather than ciliary movement, these gene could encode for proteins associated with the movement of scintillons. If this is the case then I have identified a candidate protein within *P. lunula* that is regulated by the diel cycle. A follow up experimental study would be needed to determine if this kinesin-like protein is involved in scintillon movement within *P. lunula*. Conserved domains within these genes could serve as candidate domains for further *in silico* analysis.

4.3 Pathways of interest in bioluminescence

Pathways investigated here included circadian rhythm, fluid shear stress, cAMP, and actin cytoskeleton regulation. Circadian rhythm was chosen in order to establish a connection between *P. lunula* transcripts and circadian regulation and the diel cycle. Most *P. lunula* transcripts were upregulated when mapped against the human circadian regulation pathway aside from the CLOCK gene. The CLOCK and BMAL1 proteins are transcriptional regulators helping to define circadian rhythms within an organism (Dunlap 1999). The upregulation of most transcripts that mapped to this pathway suggests the presence of transcriptional regulation of genes under dark conditions that regulate the circadian rhythm in dinoflagellates. The first report of circadian regulation at the transcript level in *P. lunula* was made by Okamoto and Hastings in a 2003 microarray analysis (Okamoto and Hastings 2003). Of the sequenced genes, it was concluded that 3% of dinoflagellate transcript levels were controlled by circadian clock

genes. Considering these essential circadian oscillators are at odds - the BMAL1 protein being upregulated and CLOCK being downregulated in the dark phase - the implications are unclear but there may be a difference in how these transcription factors regulate gene expression (i.e. upregulation vs suppression of circadian regulated genes). One study in mice revealed that BMAL1 is an essential clock gene as its deletion resulted in a loss of circadian rhythmicity (Bunger et al. 2000). In another study also in mice displayed that in certain tissues, BMAL1 may not be necessary to maintain circadian rhythmicity (Hanukoglu and Hanukoglu 2016). The competing results indicate a greater complexity to the circadian system than previous studies show. The differential gene regulation of either master clock gene has unclear implications but in this case, CLOCK could be used for regulating genes related to functions useful primarily within the light phase which could explain the downregulation in our samples during the dark phase.

In the fluid shear stress pathway, Transcripts that mapped to IL-IR, ERK5, PI3K, and β-catenin were upregulated in the dark and may contribute to the transformation of mechanical signals to chemical signals within the dinoflagellate. In humans, IL-1 cytokines are responsible for inflammatory regulation within the immune responses (Kaneko et al. 2019). The receptor, IL-1R, was initially of interest because it is a membrane bound receptor. IL-1R transports cytokines, glycoproteins, and while all receptors are, to some degree, sensitive to mechanical stimuli the likelihood of an IL-1 like-receptor in *P.lunula* being involved in bioluminescence is low.

Upregulation of kinases PI3K and ERK5 indicate cellular communication and neuronal plasticity in human cells (Tubita et al. 2021, Evangelisti and Martelli 2023). Within *P. lunula* cultures these proteins could have implications for general cellular communication and/or response to the diel cycle seeing as they are upregulated in the dark part of the cycle. Higher cellular communication during the dark period could indicate communication occurring in the cell in preparation for bioluminescent activity. PIK3 is known to produce secondary messengers followed by the activation of either receptor tyrosine kinases (RTK) or GPCRs. Kinases facilitate a wide variety of intracellular signaling, therefore, kinases should not be ruled out of future investigations into dinoflagellates bioluminescence.

β-catenin is responsible for cell to cell junctions and regulating gene transcription in humans as well as cadherin-mediated cell adhesion at the plasma membrane (Norvell et al. 2004, Brembeck et al. 2006). In general, the upregulation of these proteins indicate intracellular signaling, but may also indicate some loose association with the communication between scintillons and vacuoles. It is hypothesized that in addition to facilitating the relocation of scintillons, cytoskeletal components may also tether scintillons to the vacuolar membrane (Heimann et al. 2009). The physical association between the scintillons and vacuole is thought aid in transferring the action potential from one organelle to the other (Valiadi and Iglesias-Rodriguez 2013)

The cAMP signaling pathway, revealed upregulation of two voltage-gated channels, CNGC and VDCC, and a GPCR subunit in the dark treatment. The two voltage-gated channels are known to be involved in Ca^{2+} movement, potentially aiding in the acquisition of Ca^{2+} during

bioluminescence. The upregulation of a GPCR subunit may have implications that GPCR activity is occurring during the dark phase. Identification of GPCRs within the dataset without considering transcript abundance may be the next step in *in silico* determination of GPCR involvement in bioluminescence.

The selection of cytoskeletal regulation is relevant to bioluminescence as motor proteins are hypothesized to be responsible for facilitating movement of the scintillons, thereby enabling bioluminescence. Upregulated proteins in the actin pathway like Arp2/3, PFN, and ERM were all of interest due to their relation to actin, a protein that provides structure to a cell and plays an important role in cell migration (Ramachandran et al. 2000, Koestler et al. 2013, Solinet et al. 2013). The primary role that a motor protein would play in bioluminescence is via the movement of scintillons to the periphery of the cell. Upregulation of actin-related proteins and kinesin-like proteins provide a solid foundation for further exploration of these motor proteins *in silico* and *in vitro*. Investigations into motor proteins such as these within *P. lunula* could reveal, for the first time, specific proteins responsible for this integral part of bioluminescence in this species.

4.4 Conclusions and future directions

The transcriptomic analysis conducted during this study revealed metabolic differences in *P. lunula* under dark and light conditions where photosynthesis was generally downregulated while central carbon pathways of TCA cycle and genes involved in circadian rhythm were upregulated under dark conditions. More specifically, each of the approaches for

investigating differentially expressed genes or putative transcripts involved in bioluminescence pathways revealed up- or downregulated genes under dark conditions.

Results did not yield a clear indication of a mechanosensitive protein that would initiate bioluminescence. However, while this study did not reveal a specific protein, it does not support evidence against the receptor being a GPCR as the GPCR mRNA transcripts may not have been captured in the RNA sequencing or may be poorly annotated. It is also possible that the mechanoreceptor involved in bioluminescence is not transcriptionally regulated and thus would not necessarily be differentially expressed under dark conditions. While support for a specific GPCR is not present at this time, there is support for GPCR involvement such as the α-subunit of a G-protein as seen in the results from the cAMP signaling pathway. There is also support for downstream signaling components, as in PI3K signaling, that could function with GPCRs such as the many upregulated kinases seen in figure 6 (Evangelisti and Martelli 2023). Additionally, while not yet identified in our dataset, TRPA1 in *Drosophilia* larvae have been shown to be sensitive to shear stress (Gong et al. 2022). Considering TRP-mediated mechanotransduction, this opens up more avenues of exploration for future *in silico* work in dinoflagellate bioluminescence. For the several intramembrane channels found, a search of conserved protein domains within existing databases compared to available dinoflagellate transcriptomes and genomes could help elucidate presence of TRP channels and GPCRs. Additional research is to be done to investigate their interactions. Findings within the actin regulation pathway reveal the presence and expression of actin-related genes, the Arp2/3, PFN, and ERM, that may be involved in the movement of scintillons. This result supports the hypothesized proposed by Seo and Fritz' suggestion that motor proteins are involved in the

movement of scintillons. Fluorescence of actin could be used to determine the presence and movement of actin within the dinoflagellate cell.

Due to the possibility that the target mechanoreceptor may not be translationally regulated, there are a few methods that could be used to identify a potential mechanoreceptor. Investigation in to relevant protein domains with available genomes and transcriptomes, specifically, to other dinoflagellates species such as *Symbiodinium*. After *in silico* determination of conserved protein domains, an experiment using qPCR with genes of interest that code for motor proteins or anonymous GPCR would be interesting to move forward with as this method would allow for quantification of those genes directly for use in additional analysis. Another direction of interest would be an analysis of the literature comparing genes of interest found in this study with plant proteins in order to identify specific conserved characteristics which can then be used to also search the available dinoflagellates transcriptomes. Genes of particular interest include those that code for the G-protein α-subunit, voltage-gated ion channels, circadian transcription factors CLOCK and BMAL1, and actin and kinesin-like proteins. These methods would further elucidate dinoflagellates candidate proteins for the proposed mechanoreceptor, proteins that may be responsible for transcriptional regulation, and motor proteins. This experimental work contributes in part to the larger body of research in dinoflagellates bioluminescence. These data also provide a foundation for future work by increasing the amount of publically available sequencing data on bioluminescent dinoflagellates.

References

Afgan, E., Nekrutenko, A., Grüning, B.A., Blankenberg, D., Goecks, J., Schatz, M.C., Ostrovsky, A.E. et al. 2022. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res.* 50:W345–51.

Anderson, D.M., Fensin, E., Gobler, C.J., Hoeglund, A.E., Hubbard, K.A., Kulis, D.M., Landsberg, J.H. et al. 2021. Marine harmful algal blooms (HABs) in the United States: History, current status and future trends. *Harmful Algae*. 102:101975.

Berthiaume, C. 2013. keggannot. *Github*. https://github.com/ctberthiaume/keggannot.git

de Wet, J.R., Wood, K. V, Helinski, D.R. & DeLuca, M. 1985. Cloning of firefly luciferase cDNA and the expression of active luciferase in Escherichia coli. *Proc. Natl. Acad. Sci.* 82:7870–3.

Bhalla, V. & Hallows, K.R. 2008. Mechanisms of ENaC regulation and clinical implications. *J. Am. Soc. Nephrol.* 19:1845–54.

Brembeck, F.H., Rosário, M. & Birchmeier, W. 2006. Balancing cell adhesion and Wnt signaling, the key role of β-catenin. *Curr. Opin. Genet. Dev.* 16:51–9.

Bryk, A.A., Blagonravov, M.L., Goryachev, V.A., Chibisov, S.M., Azova, M.M. & Syatkin, S.P. 2022. Daytime exposure to blue light alters cardiovascular circadian rhythms, electrolyte excretion and melatonin production. *Pathophysiology*. 29:118–33.

Buchfink, B., Reuter, K. & Drost, H.-G. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods*. 18:366–8.

Bunger, M.K., Wilsbacher, L.D., Moran, S.M., Clendenin, C., Radcliffe, L.A., Hogenesch, J.B., Simon, M.C. et al. 2000. Mop3 is an essential component of the master circadian

pacemaker in mammals. *Cell*. 103:1009–17.

Chang, Z. & Fu, X. 2023. biogenesis of secretory proteins in eukaryotic and prokaryotic cells. *Encycl. Cell Biol.* 689–702.

Chen, A.K., Latz, M.I., Sobolewski, P. & Frangos, J.A. 2007. Evidence for the role of G-proteins in flow stimulation of dinoflagellate bioluminescence. *Am. J. Physiol. - Regul. Integr. Comp. Physiol.* 292:2020–7.

Cheng, H.C., Qi, R.Z., Paudel, H. & Zhu, H.J. 2011. Regulation and Function of Protein Kinases and Phosphatases. *Enzyme Res.* 2011.

Dunlap, J.C. 1999. Molecular bases for circadian clocks. *Cell*. 96:271–90.

Evangelisti, C. & Martelli, A.M. 2023. The PI3K/Akt/mTOR pathway. *Encycl. Cell Biol.* 153–61.

Fajardo, C., Amil-Ruiz, F., Fuentes-Almagro, C., De Donato, M., Martinez-Rodriguez, G., Escobar-Niño, A., Carrasco, R. et al. 2019. An "omic" approach to Pyrocystis lunula: New insights related with this bioluminescent dinoflagellate. *J. Proteomics*. 209.

Fajardo, C., De Donato, M., Rodulfo, H., Martinez-Rodriguez, G., Costas, B., Mancera, J.M. & Fernandez-Acero, F.J. 2020. New perspectives related to the bioluminescent system in dinoflagellates: Pyrocystis lunula, a case study. *Int J Mol Sci.* 21(5):1784.

Goldmann, W.H. 2014. Mechanosensation: A basic cellular process. *Prog. Mol. Biol. Transl. Sci.* 126:75–102.

Gomez, F. 2020. Chapter 1. Diversity and classification of dinoflagellates. In *dinoflagellates: classification, evolution, physiology and ecological significance*. Nova Science., New York.

Gong, J., Chen, J., Gu, P., Shang, Y., Ruppell, K.T., Yang, Y., Wang, F. et al. 2022. Shear

stress activates nociceptors to drive Drosophila mechanical nociception. *Neuron*. 110:3727-3742.e8.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X. et al. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* 29:644.

Hanks, S.K. & Quinn, A.M. 1991. [2] Protein kinase catalytic domain sequence database: Identification of conserved features of primary structure and classification of family members. *Academic Press.,* San Diego. pp. 38–62.

Hanukoglu, I. & Hanukoglu, A. 2016. Epithelial sodium channel (ENaC) family: Phylogeny, structure-function, tissue distribution, and associated inherited diseases HHS Public Access. *Gene*. 579:95–132.

Hardie, R.C. & Raghu, P. 2001. Visual transduction in Drosophila. *Nature*. 413:186-193.

Hastings, J. 2013. Circadian rhythms in dinoflagellates: What is the purpose of synthesis and destruction of proteins? *Microorganisms*. 1:26–32.

Heimann, K., Klerks, P.L. & Hasenstein, K.H. 2009. Involvement of actin and microtubules in regulation of bioluminescence and translocation of chloroplasts in the dinoflagellate Pyrocystis lunula. *Bot. Mar.* 52:170–7.

Hittle, B.M. & Wong, I. 2022. Blue light and sleep: What nurses need to know. *American Nurse Journal*. 17:20-23

Hong, H.-H., Lee, H.-G., Jo, J., Kim, H.M., Kim, S.-M., Park, J.Y., Jeon, C.B. et al. 2016. The exceptionally large genome of the harmful red tide dinoflagellate Cochlodinium polykrikoides Margalef (Dinophyceae): determination by flow cytometry. *Algae*. 31:373–8.

Huang, Y., Niu, B., Gao, Y., Fu, L. & Li, W. 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*. 26:680–2.

Inanobe, A. & Kurachi, Y. 2014. Membrane channels as integrators of G-protein-mediated signaling. *Biochim. Biophys. Acta - Biomembr.* 1838:521–31.

Johnson, C.H., Inoué, S., Flint, A. & Hastings, J.W. 1985. Compartmentalization of algal bioluminescence: autofluorescence of bioluminescent particles in the dinoflagellate Gonyaulax as studied with image-intensified video microscopy and flow cytometry. *J. Cell Biol.* 100:1435–46.

Kahn-Kirby, A.H., Dantzker, J.L.M., Apicella, A.J., Schafer, W.R., Browse, J., Bargmann, C.I. & Watts, J.L. 2004. Specific polyunsaturated fatty acids drive TRPV-dependent sensory signaling in vivo. *Cell*. 119:889–900.

Kanehisa, M. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28:27–30.

Kaneko, N., Kurata, M., Yamamoto, T., Morikawa, S. & Masumoto, J. 2019. The role of interleukin-1 in general pathology. *Inflamm. Regen. 2019 391*. 39:1–16.

Koestler, S.A., Steffen, A., Nemethova, M., Winterhoff, M., Luo, N., Holleboom, J.M., Krupp, J. et al. 2013. Arp2/3 complex is essential for actin network treadmilling as well as for targeting of capping protein and cofilin. *Mol. Biol. Cell*. 24:2861–75.

Konjikusic, M.J., Lee, C., Yue, Y., Shrestha, B.D., Nguimtsop, A.M., Horani, A., Brody, S. et al. 2023. Kif9 is an active kinesin motor required for ciliary beating and proximodistal patterning of motile axonemes. *J. Cell Sci.* 136.

Lang, S., Erdmann, F., Jung, M., Wagner, R., Cavalie, A. & Zimmermann, R. 2011. Channels Sec61 complexes form ubiquitous ER Ca 2+ leak channels. *Channels*.

5:21102557.

Lindström, J.B., Pierce, N.T. & Latz, M.I. 2017. Role of TRP channels in dinoflagellate mechanotransduction. *Biol. Bull.* 233:151–67.

Love, M.I., Huber, W. & Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550.

Ma, Q., Ye, L., Liu, H., Shi, Y. & Zhou, N. 2017. An overview of Ca2+ mobilization assays in GPCR drug discovery. Taylor and Francis Ltd.

Manning, A.L., Ganem, N.J., Bakhoum, S.F., Wagenbach, M., Wordeman, L. & Compton, D.A. 2007. The kinesin-13 proteins kif2a, kif2b, and kif2c/mcak have distinct roles during mitosis in human cells. *Mol. Biol. Cell*. 18:2970–9.

Martin, J., Anamika, K. & Srinivasan, N. 2010. Classification of protein kinases on the basis of both kinase and non-kinase regions. *PLoS One*. 5:1–12.

McClung, C.R. 2006. Plant circadian rhythms. *Plant Cell*. 18:792.

Mobasheri, A., Barrett-Jolley, R., Shakibaei, M., Canessa, C.M. & Martín-Vasallo, P. 2005. Enigmatic roles of the epithelial sodium channel (enac) in articular chondrocytes and osteoblasts: mechanotransduction, sodium transport or extracellular sodium sensing? *Mechanosensitivity in Cells and Tissues*. Academia.

Montell, C. 2012. Drosophila visual transduction. *Trends Neurosci.* 35:356–63.

Mulo, P., Sicora, C. & Aro, E.-M. 2009. Cyanobacterial psbA gene family: optimization of oxygenic photosynthesis. *Cell. Mol. Life Sci.* 66:3697–710.

Norvell, S.M., Alvarez, M., Bidwell, J.P. & Pavalko, F.M. 2004. Fluid shear stress induces beta-catenin signaling in osteoblasts. *Calcif. Tissue Int.* 75:396–404.

Okamoto, O.K. & Hastings, J.W. 2003. Novel dinoflagellate clock-related genes identified

through microarray analysis. *J. Phycol.* 39:519–26.

Paluch, E.K., Nelson, C.M., Biais, N., Fabry, B., Moeller, J., Pruitt, B.L., Wollnik, C. et al. 2015. Mechanotransduction: Use the force(s). *BMC Biol.* 13:1–14.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. & Kingsford, C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods*. 14:417–9.

Pfeil, E.M., Brands, J., Merten, N., Vögtle, T., Vescovo, M., Rick, U., Albrecht, I.M. et al. 2020. Heterotrimeric g protein subunit gαq is a master switch for gβγ-mediated calcium mobilization by Gi-coupled GPCRs. *Mol. Cell*. 80:940-954.e6.

Pozdnyakov, I., Matantseva, O. & Skarlato, S. 2021. Consensus channelome of dinoflagellates revealed by transcriptomic analysis sheds light on their physiology. *Algae*. 36:315–26.

Purves D, Augustine GJ, Fitzpatrick D, et al., E. 2001. Glutamate receptors. In *Neuroscience*. Sinauer Associates. 2:129-165

Ramachandran, S., Christensen, H.E.M., Ishimaru, Y., Dong, C.-H., Chao-Ming, W., Cleary, A.L. & Chua, N.-H. 2000. Profilin plays a role in cell elongation, cell shape maintenance, and flowering in arabidopsis. *Plant Physiol.* 124:1637–47.

Rodriguez, J.D., Haq, S., Bachvaroff, T., Nowak, K.F., Nowak, S.J., Morgan, D., Cherny, V. V. et al. 2017. Identification of a vacuolar proton channel that triggers the bioluminescent flash in dinoflagellates. *PLoS One*. 12:1–24.

Rosenbaum, D.M., Rasmussen, S.G.F. & Kobilka, B.K. 2009. The structure and function of G-protein-coupled receptors. *Nature*. 459:356–363.

Seo, K.S. & Fritz, L. 2000. Cell ultrastructural changes correlate with circadian rhythm in Pyrocystis lunula (Pyrrophyta). *J. Phycol*. 36:351–8.

Serin, Y. & Acar Tek, N. 2019. Effect of circadian rhythm on metabolic processes and the regulation of energy balance. *Ann. Nutr. Metab.* 74:322–30.

Shalaeva, D.N., Galperin, M.Y. & Mulkidjanian, A.Y. 2015. Eukaryotic G protein-coupled receptors as descendants of prokaryotic sodium-translocating rhodopsins. *Biol. Direct*. 10:63.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E.M. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 31:3210–2.

Solinet, S., Mahmud, K., Stewman, S.F., Ben El Kadhi, K., Decelle, B., Talje, L., Ma, A. et al. 2013. The actin-binding ERM protein moesin binds to and stabilizes microtubules at the cell cortex. *J. Cell Biol.* 202:251–60.

Tubita, A., Tusa, I. & Rovida, E. 2021. Playing the whack-a-mole game: erk5 activation emerges among the resistance mechanisms to RAF-MEK1/2-ERK1/2- targeted therapy. *Front. Cell Dev. Biol.* 9.

Valiadi, M. & Iglesias-Rodriguez, D. 2013. Understanding bioluminescence in dinoflagellates—how far have we come? *Microorganisms*. 1:3–25.

Vauquelin, G. & von Mentzer, B. 2007. *G Protein-Coupled Receptors*. John Wiley & Sons, Ltd, Chichester, UK. 252 pp.

Von Dassow, P. & Latz, M.I. 2002. The role of Ca(2+) in stimulated bioluminescence of the dinoflagellate Lingulodinium polyedrum. *J. Exp. Biol.* 205:2971–86.

Wang, M.Y. & Liu, Y.J. 2020. Mechanistic insight into initiation of dinoflagellate bioluminescence. *J. Photochem. Photobiol. A Chem.* 394.

Wang, Y., Jensen, L., Højrup, P. & Morse, D. 2005. Synthesis and degradation of

dinoflagelatte plastid-encoded psbA proteins are light-regulated, not circadian-regulated. *Proc. Natl. Acad. Sci. U. S. A.* 102:2844–9.

Whitehouse, L. (Lswhiteh) 2019. Diatom metatranscriptomics workflow. https://github.com/Lswhiteh/diatom-metatranscriptomics (last accessed March 17, 2023).

Widder, E.A. 2010. Bioluminescence in the ocean: Origins of biological, chemical, and ecological diversity. *Science*. 328:704-708

Wilson, T. & Hastings, J.W. 2013. *Bioluminescence*. Cambridge. 45–59 pp.

Tables and Figures

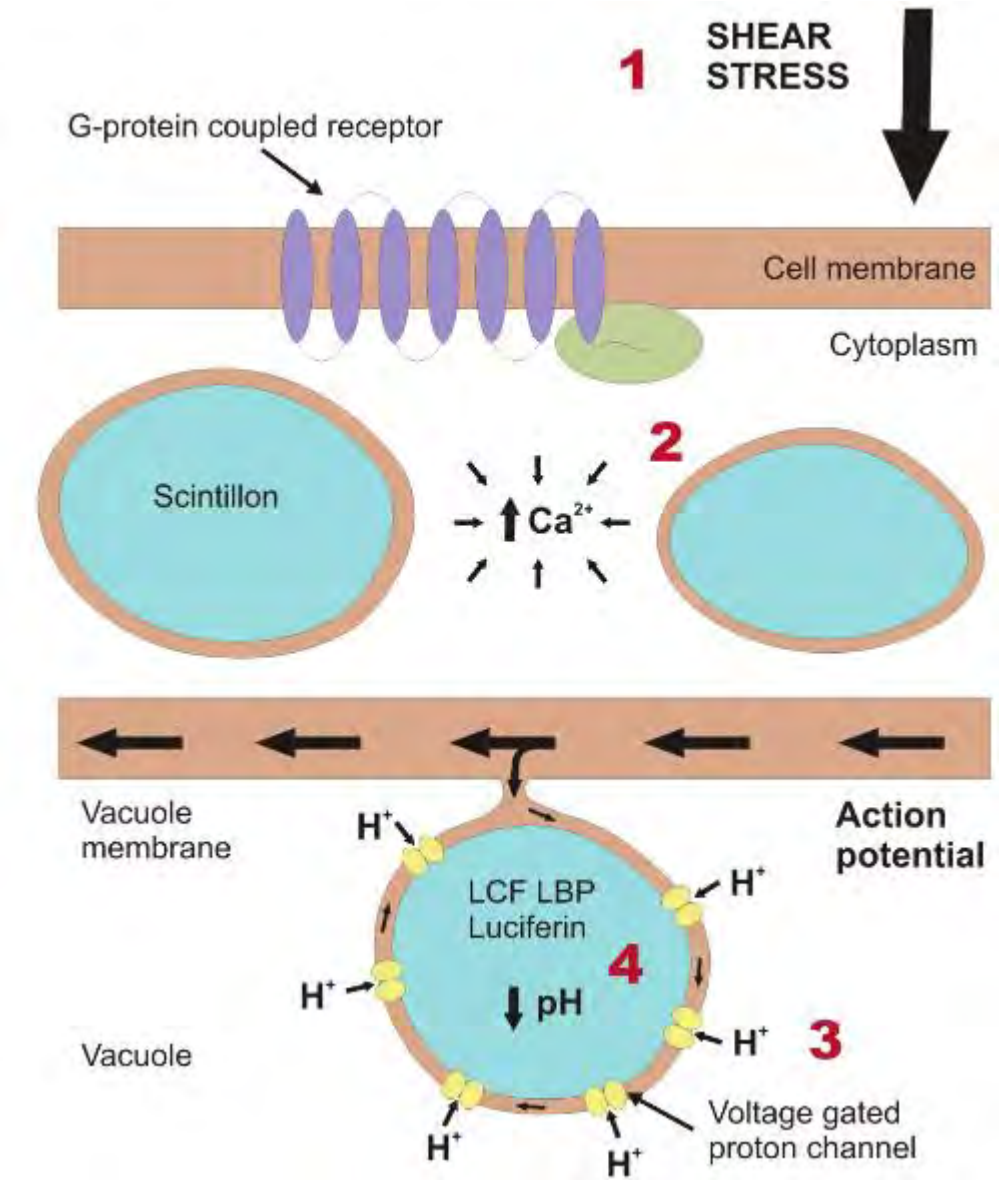

Figure 1: Schematic of a partial dinoflagellate cell illustrating cellular processes that contribute to a bioluminescent flash. 1) Mechanical force applied to the dinoflagellate's cellular membrane, triggering a transmembrane receptor (proposed to be a GPCR). 2) Intracellular Ca^{2+} concentration increases, allowing for electric signaling to occur across the vacuolar membrane. As a scintillon (containing components necessary for bioluminescence) projects into the vacuole, the generated action potential travels across the scintillon membrane. 3) As a result, voltage gated proton channels are opened. 4) This facilitates the acidification of the scintillon and through oxidation of luciferin produces bioluminescence (Valiadi and Iglesias-Rodriguez 2013).

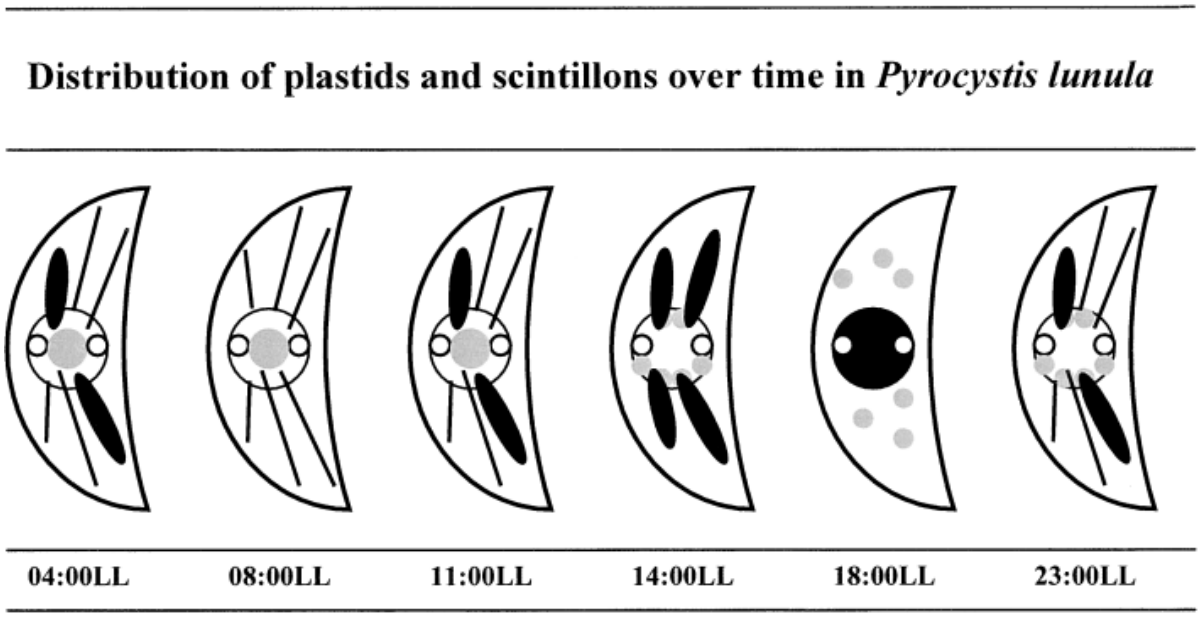

Figure 2. Schematic of morphological changes in plastids and scintillons in *P. lunula* in continuous dim light. The white circle represents the nucleus, grey circles represent scintillon distribution, and the black ovals and lines represent plastid shape (Seo and Fritz 2000).

Figure 3. Culture set up with two chambers, separated by black plastic material, set to an alternating light cycle. Flasks elevated on boxes to achieve light intensity in Bigelow's recommended range (40 -120 µEinsteins)

Table 1. Number of RNAseq reads and summary statistics of reads pre-re- and post-quality control.

Avg Raw Seqeuence Count	Avg Post - Trimmomatic Sequence Count	Sequence Length	%GC Range - Raw	%GC Range - Trimmed	Mean Quality Score
46,527,360	38,907,891	150	59-61	60-62	37.03

Figure 4. Taxonomy summary of a representative, random sample of reads. Over 90% map to Eukaryota (red) (72% Dinophyceae (light red) and the remaining eukaryotic matches (orange). The remaining percentage (<10%) in green mapped to prokaryotes.

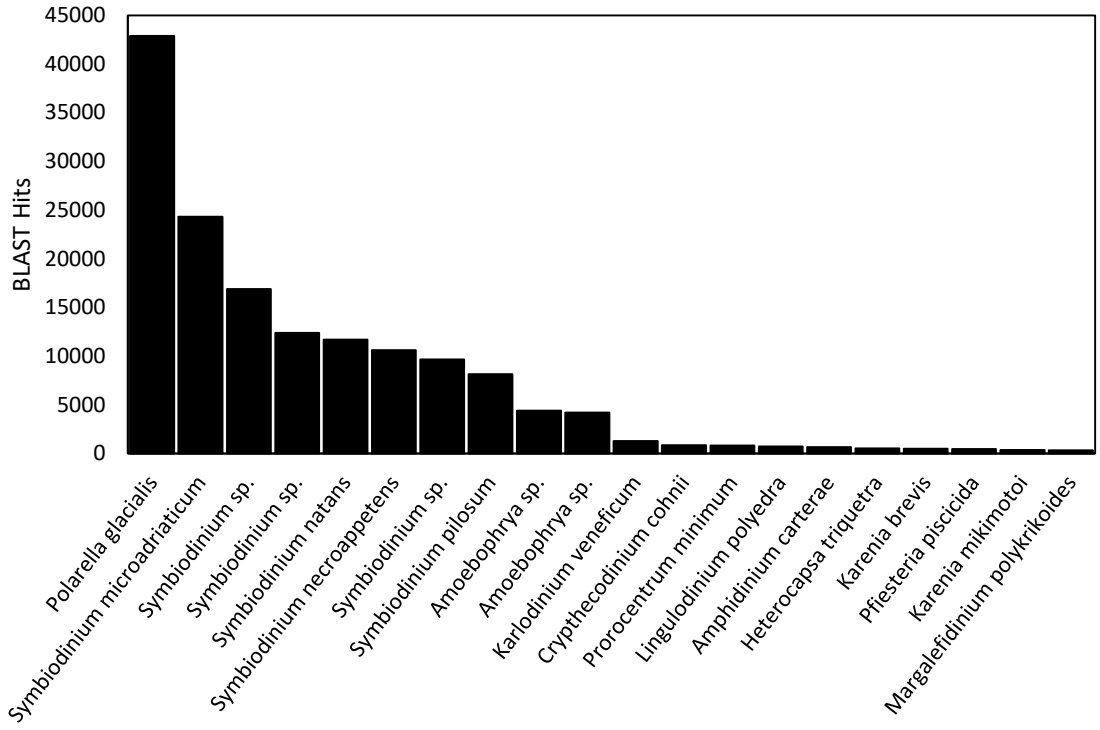

Figure 5. Top 20 organisms that mapped to the total clustered assembly (96.5%) post transcript filtering for only eukaryotes, all of which are classified as dinoflagellate or microalgae.

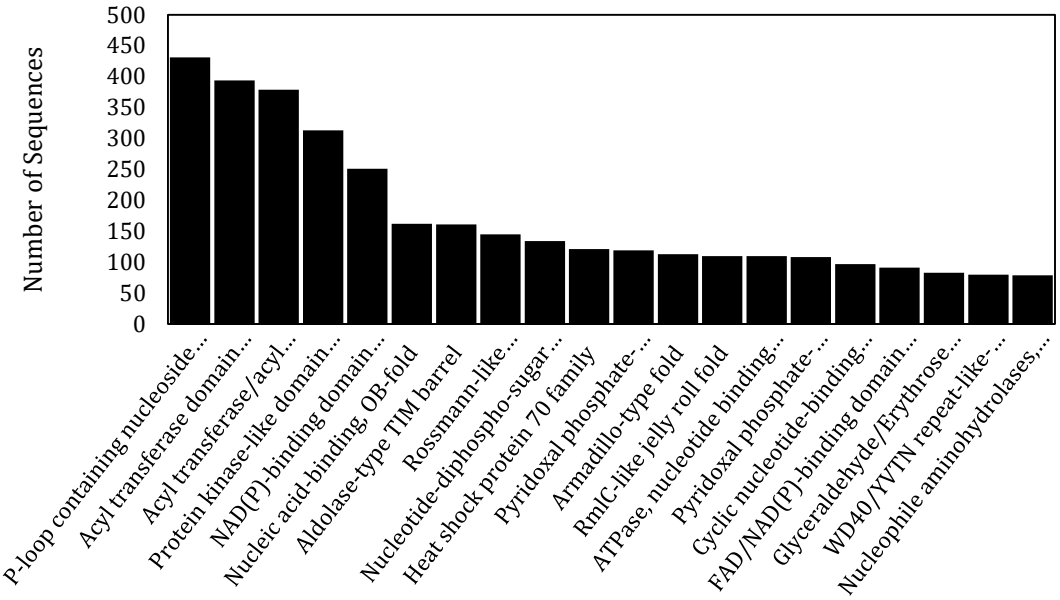

Figure 6. Top 20 protein domains generated from Interpro Scan results (22.3% of total hits), most of which were identified to be protein kinases. The top three unique domains include p-loop containing nucleosides, acyl transferase domains, and protein kinase-like domains.

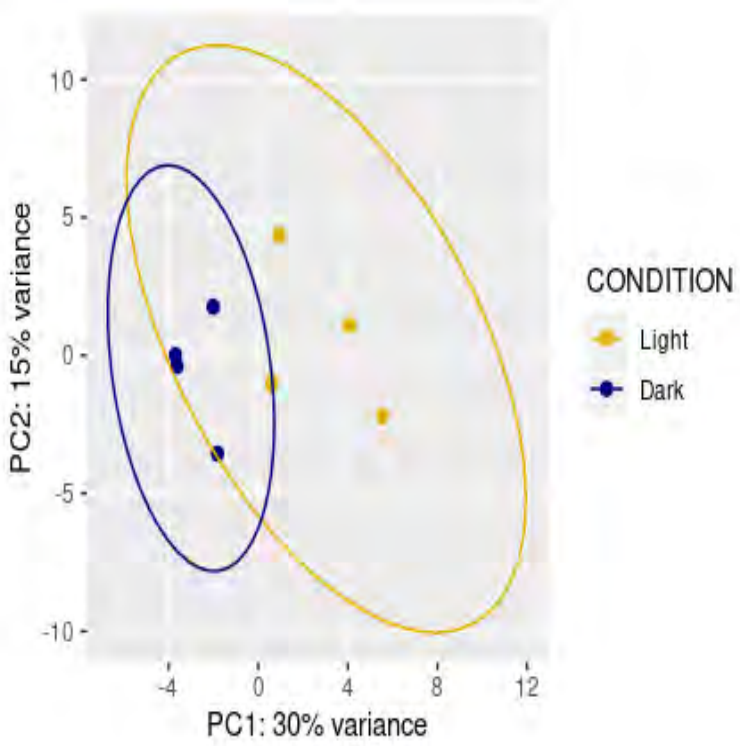

Figure 7. Principle components analysis of *P. lunula* transcriptomes. Colors represent samples from each light (yellow) or dark (blue) treatment. The X-axis represents PC1, with a 30% variance and the Y-axis displays PC2, with a 15% variance. The effect of treatment was significant (PERMANOVA, $F_{6,7} = 1.69$, $p = 0.031$).

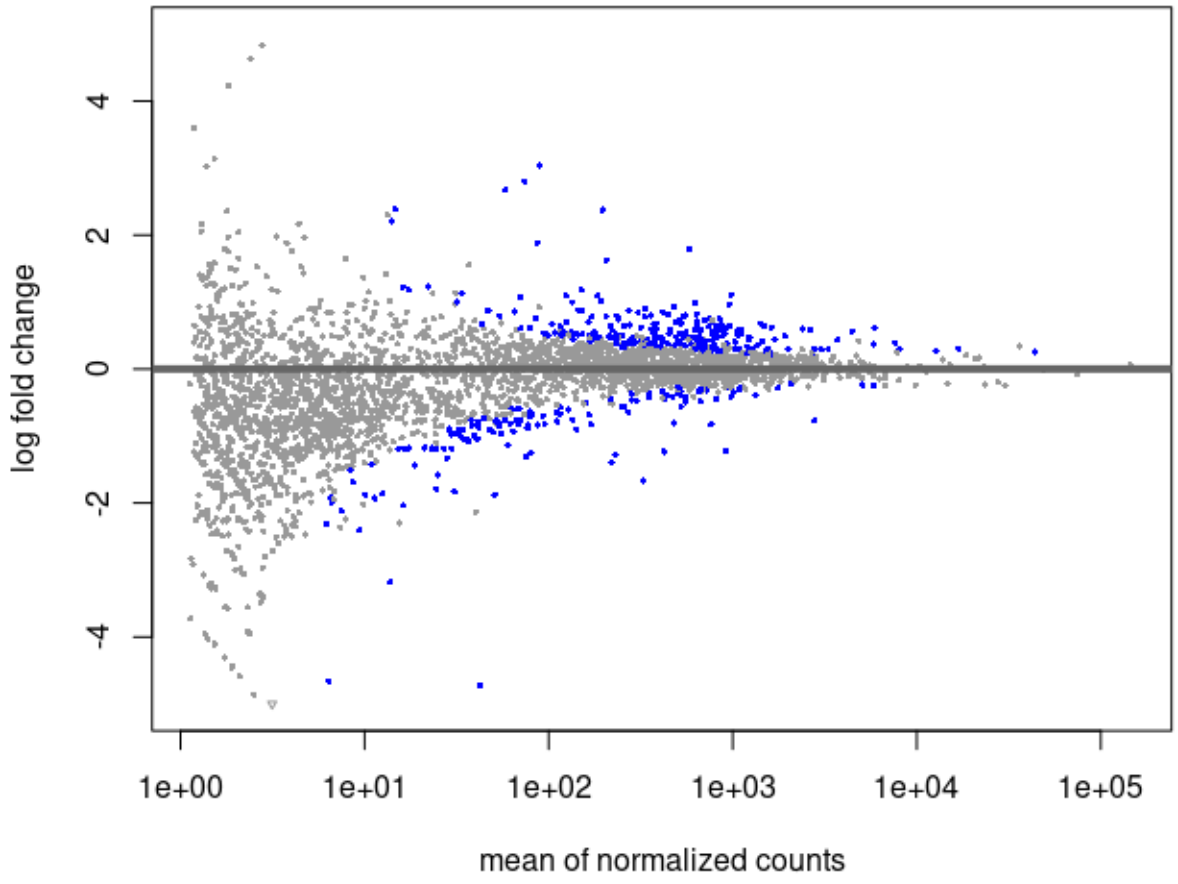

Figure 8. MA plot using a 0.05 significance threshold for contigs annotated using phegg database and remapped to the clustered assembly for abundance estimates. Blue triangles represent transcripts with a log fold change that are considered 'upregulated', above the x-axis, or 'down-regulated', below the x-axis.

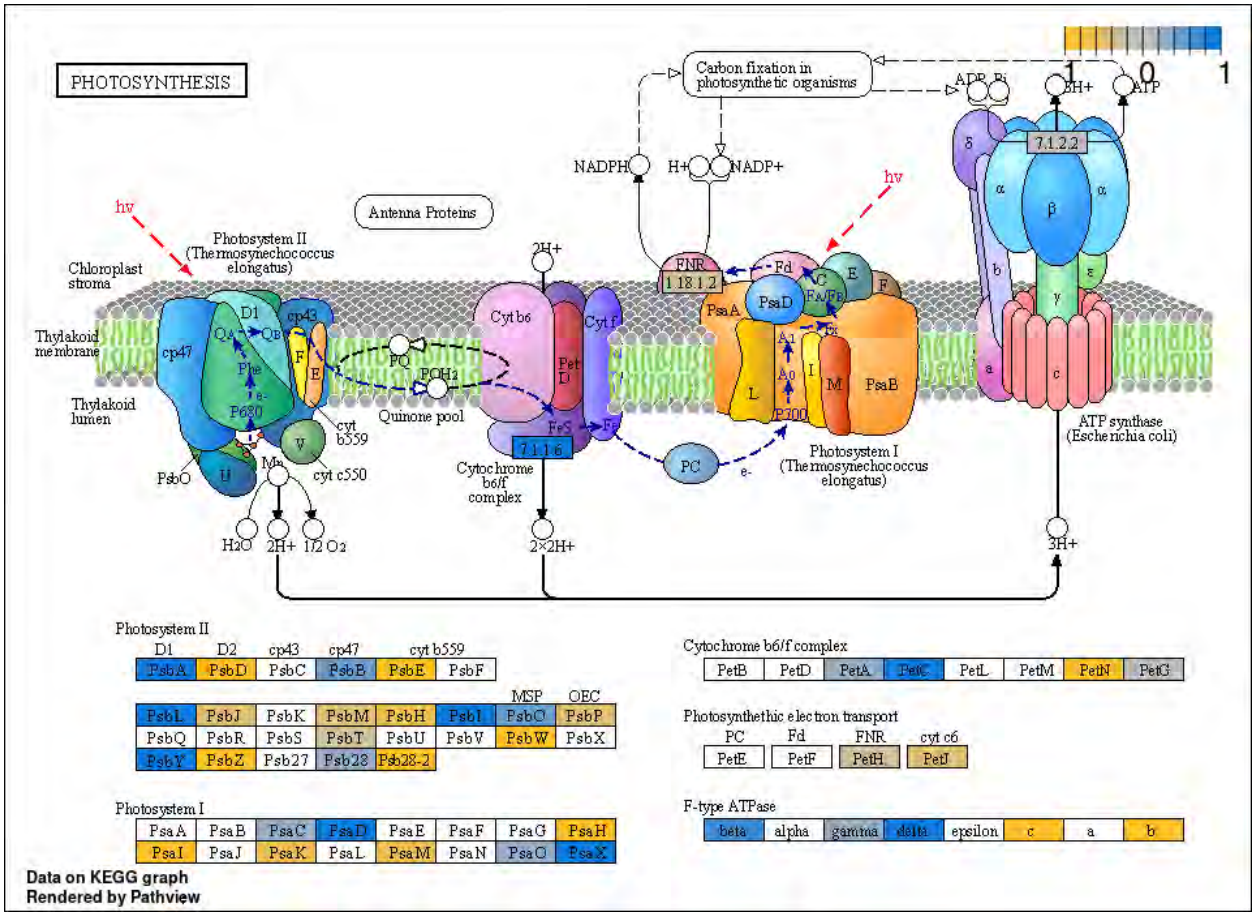

Figure 9. Pathview rendering of photosynthesis pathway (ko00195) showing relative gene expression using LFC for abundance and KO identifiers. The color scale yellow to blue to indicate down- or up-regulation of proteins, respectively.

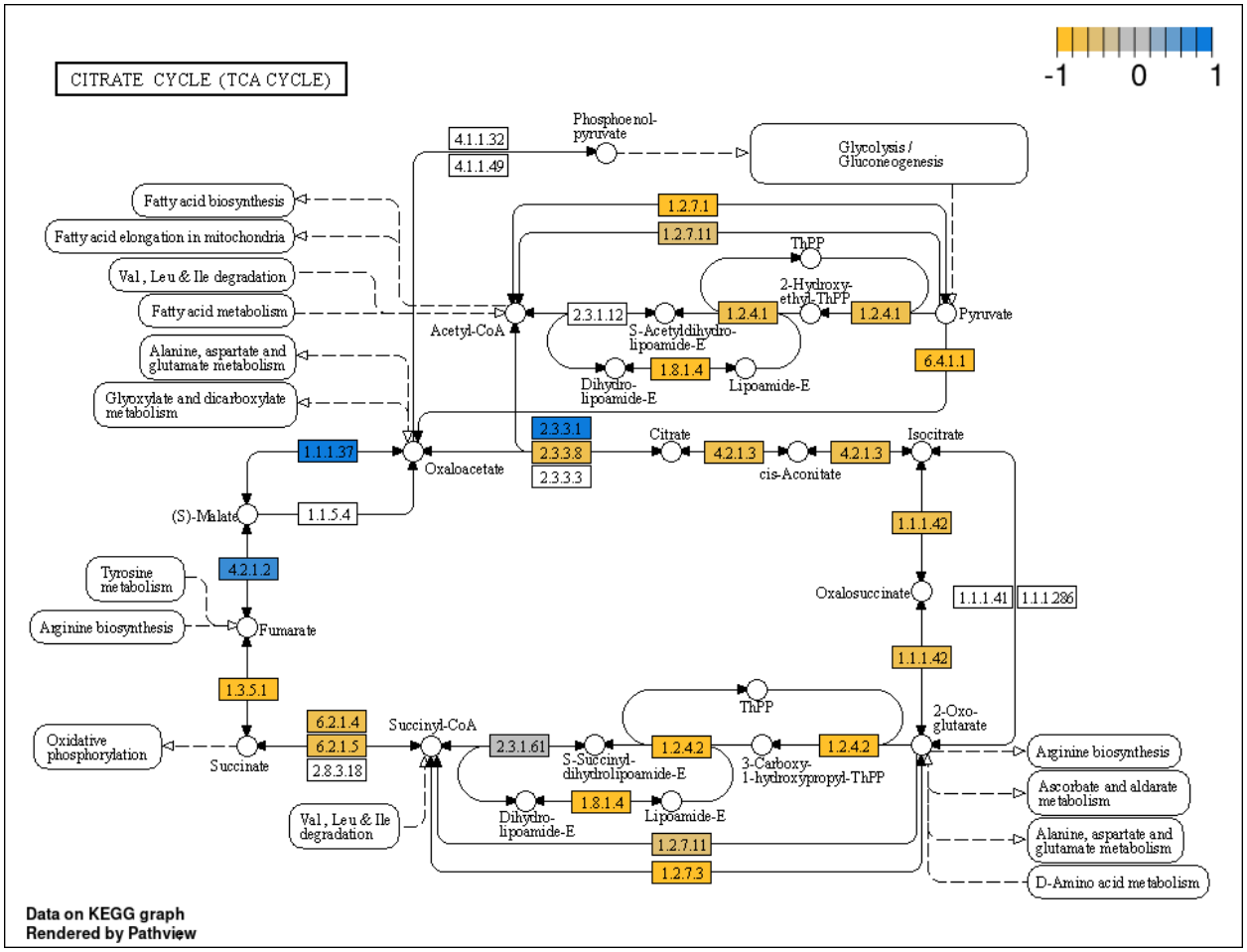

Figure 10. Pathview rendering of the TCA cycle pathway (ko00020) showing relative gene expression using LFC for abundance and KO identifiers. The color scale yellow to blue to indicate down- or up-regulation of proteins, respectively.

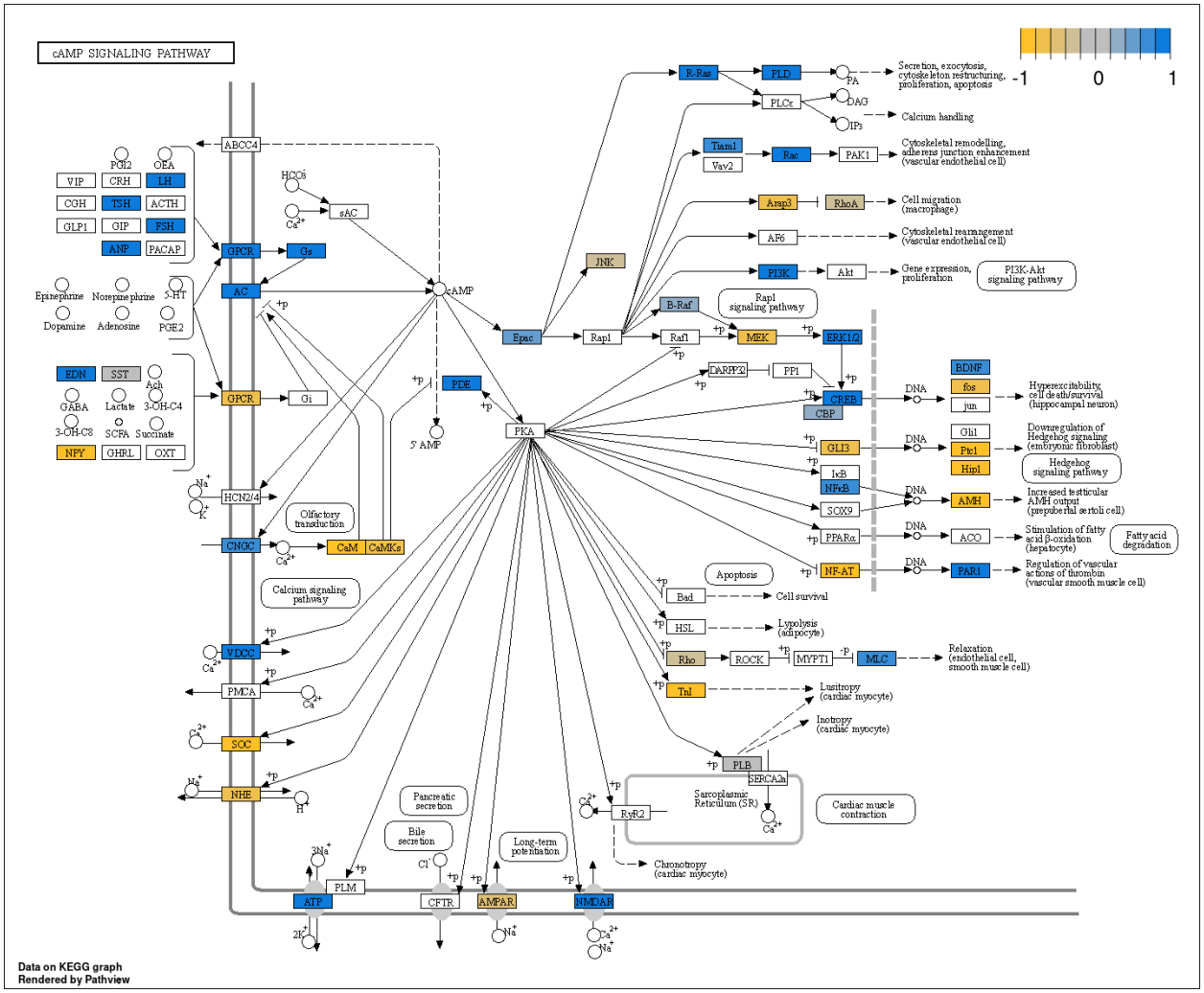

Figure 11. Pathview rendering of the cAMP signaling pathway (ko04024) showing relative gene expression using LFC for abundance and KO identifiers. The color scale yellow to blue to indicate down- or up-regulation of proteins, respectively.

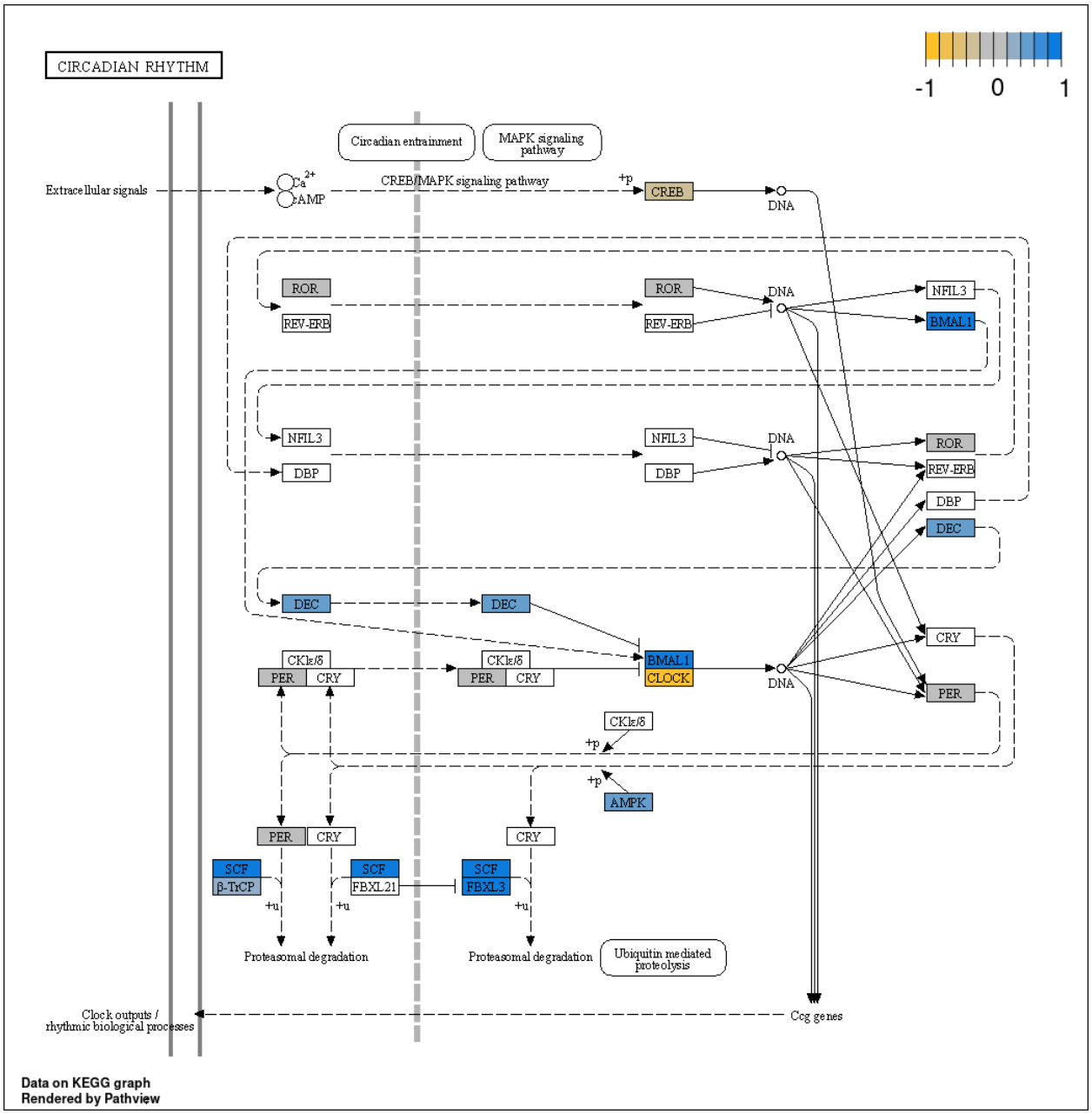

Figure 12. Pathview rendering of the circadian rhythm pathway (ko04710) showing relative gene expression using LFC for abundance and KO identifiers. The color scale yellow to blue to indicate down- or up-regulation of proteins, respectively.

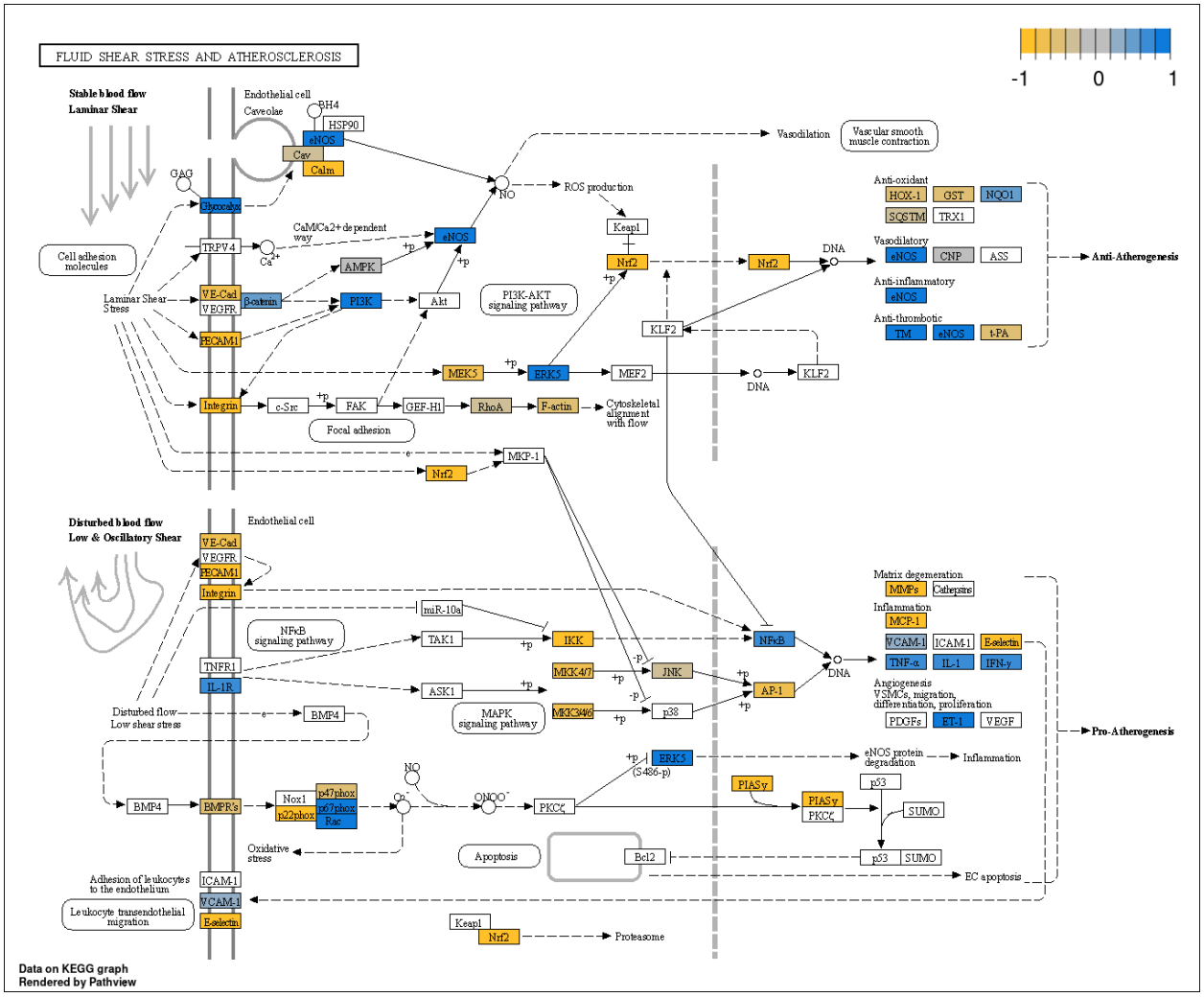

Figure 13. Pathview rendering of the fluid shear stress pathway (ko05418) showing relative gene expression using LFC for abundance and KO identifiers. The color scale yellow to blue to indicate down- or up-regulation of proteins, respectively.

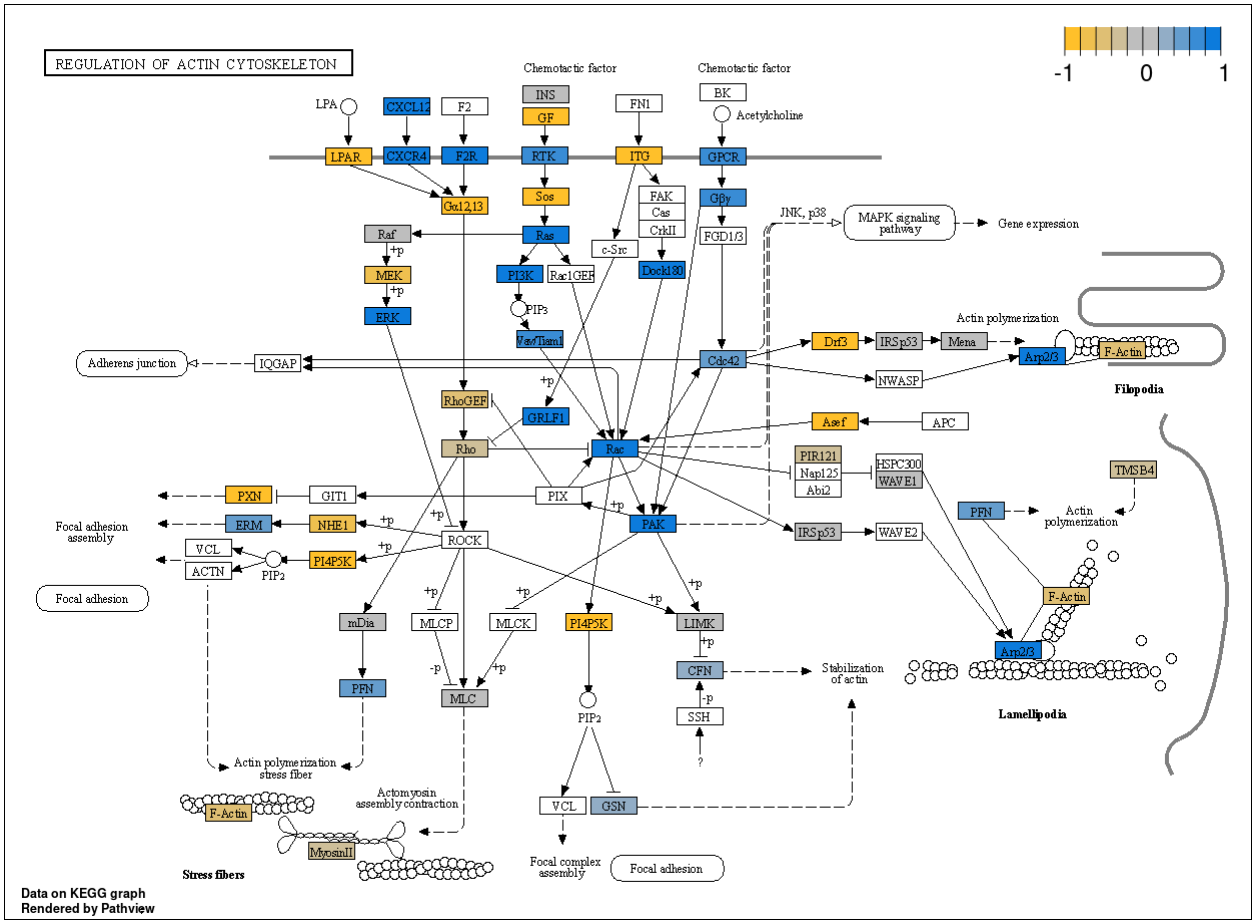

Figure 14. Pathview rendering of the actin cytoskeleton regulation pathway (ko04810) showing relative gene expression using LFC for abundance and KO identifiers. The color scale yellow to blue to indicate down- or up-regulation of proteins, respectively.

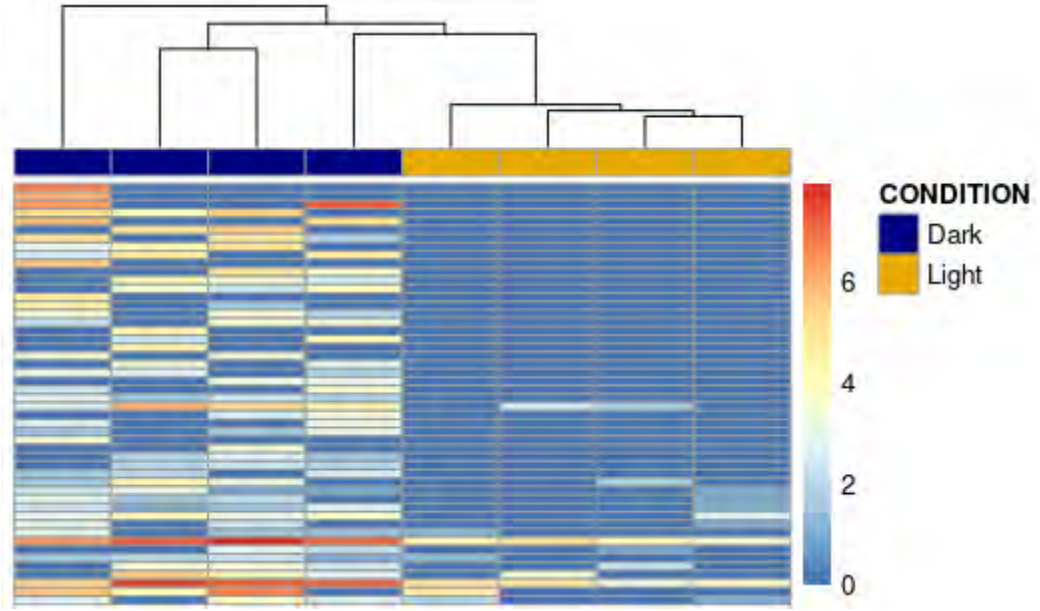

Figure 15. Phegg heatmap displaying expression patterns of differentially expressed genes between light (yellow) and dark (blue) samples. Displayed are the top 50 transcripts with the highest difference based on LFC. The gradient from dark blue to bright red indicates lowest LFC difference to highest LFC difference, respectively.

Table 2. Transcripts identified using the phegg heatmap (Fig. 15), which were compared to the NCBI nr database.

Tax	Transcript ID	Protein
Zea Maize	TRINITY_DN13150_c0_g2_i11	K10808 ribonucleoside-diphosphate reductase subunit M2
Vibrio atlanticus	TRINITY_DN40695_c0_g1_i2	3-oxoacyl-[acyl-carrier-protein] reductase K00059 3-oxoacyl-[acyl-carrier protein] reductase
Gloeophyllum trabeum	TRINITY_DN6314_c1_g1_i1	ribosomal protein s3
Sandaracinus amylolyticus	TRINITY_DN36203_c0_g1_i7 TRINITY_DN36203_c0_g1_i12 TRINITY_DN36203_c0_g1_i13 TRINITY_DN75185_c0_g1_i1	3-oxoacyl-[acyl-carrier-protein] synthase, KASII
Macaca mulatta	TRINITY_DN46767_c1_g1_i1	charged multivesicular body protein 1a
Sandaracinus amylolyticus	TRINITY_DN52559_c0_g1_i13 TRINITY_DN52559_c0_g1_i2 TRINITY_DN52559_c0_g1_i6	Bacterial, ADP-forming succinate--CoA ligase subunit β Succinyl-CoA ligase [ADP-forming] β chain

		K01903 succinyl-CoA synthetase β subunit
Breviolum minutum	TRINITY_DN102535_c0_g1_i1 TRINITY_DN16066_c0_g1_i1 TRINITY_DN16066_c0_g1_i2 TRINITY_DN16066_c0_g2_i1 TRINITY_DN16066_c0_g2_i2	transport protein SEC61 subunit β
Nicotiana tomentosiformis	TRINITY_DN44158_c0_g1_i1	photosystem I P700 chlorophyll a apoprotein A2 *P. lunula* Hit, photosystem I P700 chlorophyll a apoprotein
Sandaracinus amylolyticus	TRINITY_DN36660_c0_g1_i7	Preprotein translocase secY subunit
Breviolum minutum	TRINITY_DN5983_c0_g2_i2 TRINITY_DN5983_c0_g2_i3 TRINITY_DN5983_c0_g2_i4 TRINITY_DN5983_c0_g3_i1 TRINITY_DN5983_c1_g1_i1	ATP Synthase Subunit
Breviolum minutum	TRINITY_DN42322_c0_g1_i1	Osmotic avoidance abnormal protein 3 in non-dinoflagellate spp. KIF13B - kinesin-like protein in Symbidinium Chorismate mutase in Symbidinium ARO7/Chorismate mutase in Symbidinium
Macaca fascicularis	TRINITY_DN114012_c0_g1_i1	glycerol-3-phosphate phosphatase phosphoglycolate phosphatase

Vita

Lila Lynnette Barrera was born in Roswell, New Mexico to Theresa Annette Morley-Barrera and Cruz Muñiz Barrera. She graduated with a Bachelor of Science in Cell and Molecular Biology from Appalachian State University in December 2019. In the fall of 2020, she began her pursuit of a Master of Science in cell and molecular biology from Appalachian State University and was awarded the degree in May of 2023.