A Transcriptomics-Based Approach to Novel Photoprotein Analysis in Pelagic Molluscs

Taylor Hersh, Carnegie Mellon University

Mentor: Dr. Steven H. D. Haddock

Summer 2014

Keywords: Pelagic molluscs, Symplectin, Photoproteins, Luciferases, Bioluminescence

ABSTRACT

Bioluminescence is the production and emission of light by a living organism. While the majority of species are incapable of bioluminescence, most phyla tend to have at least one bioluminescent member. Of the different phyla, Mollusca falls only behind Chordata and Arthropoda in approximate number of bioluminescent genera (Bjorn 2008). Overall, however, the genetics and molecular biology behind bioluminescence in many different molluscs remain largely uncharacterized. This study set out to determine the primary sequence of the photoproteins/luciferases responsible for bioluminescence in six different pelagic mollusc species (*Phylliroe bucephalum, Chiroteuthis calyx, Pterygioteuthis hoylei, Vampyroteuthis infernalis, Dosidicus gigas,* and *Octopoteuthis deletron).* A known photoprotein (Symplectin) from *Sthenoteuthis oualaniensis* was used as a reference throughout this study. The Symplectin sequence was BLASTed against transcriptomes for each of the six aforementioned species. Promising alignment hits were returned for *P. bucephalum, P. hoylei, V. infernalis,* and *D. gigas*. Primers were then designed to target Symplectin-like sequences in those four species. The results suggest that *P. hoylei* contains a photoprotein that shares strong molecular similarities with Symplectin. Future studies aim to isolate the entire *P. hoylei* photoprotein/luciferase transcript and express the photoprotein/luciferase in *E. coli* cells.

INTRODUCTION

The deep sea is characterized by low light levels, patchy food sources, and extreme temperatures. To handle the challenges of living in this harsh environment, different species have developed unique adaptations that allow them to carry out the functions required for survival. Once such adaptation that allows many species to function in the deep sea is bioluminescence. Bioluminescence is the production and emission of light by a living organism. During bioluminescence, light is generated by the release of energy through a chemical reaction. Generally, this chemical reaction involves the oxidation of a light-emitting molecule, called a luciferin. The rate of oxidation of the luciferin is controlled by a luciferase or a photoprotein, both of which act as catalytic enzymes. In the case of photoproteins, the factors required for bioluminescence are bound as a unit. When an additional required ion or cofactor binds to the photoprotein, a conformational change occurs, thereby catalyzing the reaction and producing light (Haddock et al. 2009).

Across many marine and terrestrial organisms, luciferins tend to be highly conserved. For example, the luciferin coelenterazine is used in at least nine different phyla (Haddock et. al, 2009). In contrast, luciferases and photoproteins are much more varied across species. Among squids alone there are at least 70 luminous genera (Herring 1987), the majority of which have uncharacterized photoproteins/luciferases. Different squid species have been documented using bioluminescence in both defensive and offensive ways. For example, the deep-sea squid *Octopoteuthis deletron* has been seen detaching bioluminescent arm tips when threatened (Bush et al. 2009). It is also thought that certain species of squid use their bioluminescent capabilities to lure, stun, or confuse prey (Haddock et al. 2009). While behavioral records of bioluminescence in squid are becoming more plentiful, relatively little is known about the molecular biology and genetics responsible for these bioluminescent displays.

Perhaps the best characterized example of bioluminescence in squid is that of *Watasenia scintillans*, the firefly squid. *W. scintillans* emits blue light (470 nm) from many photophores spread across its mantle and head. This squid also has light organs on the tip of its fourth pair of ventral arms. This particular bioluminescence reaction involves magnesium ions, ATP, coelenterazine-disulfate (the luciferin), and a membrane-bound luciferase. When modifications were made to the coelenterazine-disulfate structure, the light-emitting capabilities were greatly diminished. Additionally, researchers found that the luciferase present in the arm light organs of *W. scintillans* is stored in cellular particles (Teranishi & Shimomura 2008).

Another well-characterized bioluminescent squid species is the purpleback flying squid,

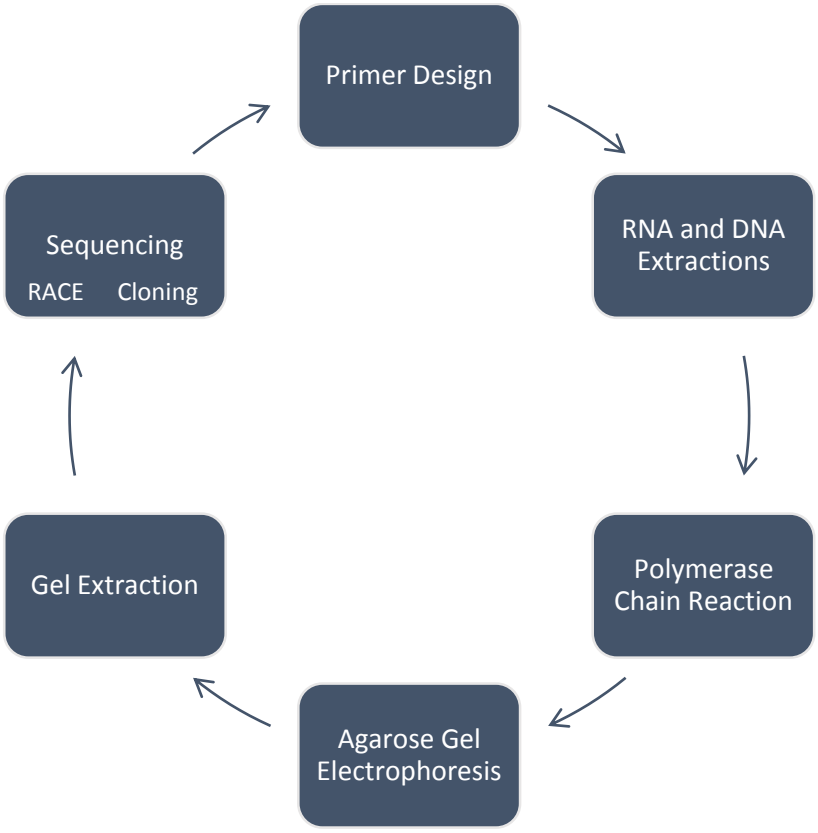
Sthenoteuthis oualaniensis (previously known as *Symplectoteuthis oualaniensis*). This squid was the subject of numerous bioluminescence studies beginning as early as 1965 and continuing today (Clarke 1965). *S. oualaniensis* is unique in that it utilizes a naturally occurring chemical variant of coelenterazine—dehydro-coelenterazine—as its luciferin (Takahasi & Isobe 1993). Additionally, while most other thoroughly researched bioluminescence reactions use divalent cations as a trigger, the bioluminescence in *S. oualaniensis* is uniquely triggered by monovalent cations (Fujii et al. 2002). In 2002, scientists were able to isolate a 60 kDa photoprotein from small mantle granules on *S. oualaniensis* (Fujii et al. 2002). This photoprotein, named Symplectin, emits a blue light when stimulated. Through subsequent analysis and experimentation, the researchers were able to determine the protein sequence of Symplectin, thereby achieving the remarkable feat of elucidating the first squid photoprotein sequence. They found that the peptide sequence for Symplectin differed from other known photoproteins but shared similarities with carbon-nitrogen hydrolases.

In the present study, the Symplectin sequence was used as a query to search through novel transcriptomic databases for six pelagic mollusc species: one nudibranch (*Phylliroe bucephalum)* and five cephalopods (*Chiroteuthis calyx, Pterygioteuthis hoylei, Vampyroteuthis infernalis, Dosidicus gigas,* and *Octopoteuthis deletron*). The purpose of this study was to find novel photoproteins/luciferases in these pelagic molluscs and to evaluate sequence similarity with Symplectin. In doing so, we hoped to further characterize the bioluminescence systems of a variety of pelagic molluscs and contribute to a greater understanding of bioluminescence in the deep-sea.

MATERIALS AND METHODS

A cyclical methodology (summarized in Figure 1) was employed in this experiment.

Figure 1: A graphical depiction of the methodological cycle that was used in this experiment.



Initially, primers were designed using the Symplectin protein sequence and the transcriptomes for each species. Those primers were then tested during PCR on extracted genomic DNA from each species. PCR products were run on 1% agarose gels, and promising bands were cut from the gel and sequenced. Sequencing results were then used to design additional primers, and the cycle was repeated. Cloning and transformation as well as rapid amplification of cDNA ends (RACE) were used to achieve better sequencing results as well. A more detailed discussion of each facet of the cycle can be found below.

Primer Design

The protein sequence for Symplectin was accessed through UniProtKB/Swiss-Prot (accession: C6KYS2). It was then BLASTed (using an in-house transcriptome database) against transcriptomes for *P. hoylei, D. gigas, V. infernalis, P. bucephalum, C. calyx,* and *O. deletron*. The transcriptomes were generated and processed per Francis et al. (2013). The best hits returned for each species were aligned to

the Symplectin protein sequence using the Seaview program. Phylogenetic trees were also constructed in order to examine potential evolutionary relationships among the species. Using these hits, both sequencing and PCR primers were designed using the PrimerQuest tool (Integrated DNA Technologies) and the Geneious software (Biomatters: Version 7.1.5).

Specimen Collection

Tissues thought to contain photoproteins/ luciferases for *P. hoylei, D. gigas, V. infernalis,* and *P. bucephalum* were collected on a variety of sea expeditions spanning 2009-2012. The acquisition information for each tissue can be seen in Table 1.

Table 1: This table summarizes the circumstances of collection for each tissue that was used in this study.

Species	**Date Collected**	**Method of Collection**	**Location of Collection**
Pterygioteuthis hoylei	02/25/12	Trawl	Gulf of California
Dosidicus gigas	06/21/10	Jigging	Gulf of California
Vampyroteuthis infernalis	12/11/09	Remotely Operated Vehicle (ROV)	Monterey Bay
Phylliroe bucephalum	06/06/10	Blue-water dive	Gulf of California

RNA and DNA Extractions

RNA and DNA extractions were carried out using tissue samples from *P. hoylei, D. gigas, V. infernalis,* and *P. bucephalum.* An AllPrep DNA/RNA/Protein Mini Kit (QIAGEN: 80004) was used for these extractions. Tissues were stored in liquid nitrogen prior to being ground with a mortar and pestle and further homogenized using a needle and syringe. For each extraction, 20-30 mg of tissue was homogenized in 600 µl of Buffer RLT. The optional DNase digestion was performed for each RNA extraction as well. RNA and DNA concentrations were quantified using a Qubit 2.0 Fluorometer (Life Technologies: Q32866). Specifically, the Qubit double-stranded DNA Broad Range Assay and the Qubit RNA High Sensitivity Assay kits were used. All work areas and tools were sterilized using RNase-Zap (Life Technologies: AM9780).

Polymerase Chain Reaction (PCR)

Extracted genomic DNA was amplified through PCR using the Phusion High-Fidelity PCR Kit (New England Biolabs: EO553S). The reaction mixture consisted of 2.5 µl forward primer, 2.5 µl reverse primer, 5 µl 10X BSA, 1 µl template DNA, 14 µl nuclease-free water, and 25 µl Phusion Master Mix. The cycling conditions were as follows: 25 cycles of 98°C for 5 sec, 50°C/temperature gradient for 10 sec, and 72°C for 60 sec. A Veriti 96-Well Thermal Cycler (Applied Biosystems: 4375305) was used for all PCR reactions.

Agarose Gel Electrophoresis and Gel Extraction

PCR products were analyzed using 1% agarose gels. Gels were run at 60-80 volts in 1X TAE buffer. Gel imaging was carried out using the ImageJ program. When promising bands were seen on a gel, they were excised using a sterile razor blade. The QIAquick Gel Extraction Kit (QIAGEN: 28704) was used to isolate DNA from the bands.

Sequencing

To prepare the products for sequencing, the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems: 4337455) was used. The sequencing reaction mixture consisted of 0.5 µl Big Dye, 1.75 µl 5X Sequencing Buffer, 0.5 µl primer (3.2 pmol/µl), 1 µl DNA, and 6.25 µl nuclease-free water. Ninety-six-well plates were used for sequencing reactions. The cycling conditions were as follows: 25 cycles of 96°C for 10 sec, 50°C for 5 sec, and 60°C for 4 min. The reaction mixture for precipitating the DNA consisted of 1 µl 125 mM EDTA, 1 µl 3 M sodium acetate, and 25 µl 95% ethanol. 10 µl of HiDi formamide was added to each well prior to sequencing, and sequencing was carried out by Lynne Christianson of the Monterey Bay Aquarium Research Institute (MBARI). Sequencing results were analyzed using the Geneious software.

Cloning and Transformation

Promising PCR products were cloned into One Shot chemically competent *E. coli* cells using the Zero Blunt TOPO PCR Cloning Kit for Sequencing (Invitrogen: 1219893). The reaction components included 1 µl PCR product, 1 µl salt solution, 3 µl nuclease-free water, and 1 µl TOPO vector. The reactions were incubated for 20 minutes at room temperature because they had low DNA concentrations.

Products were plated on LB plates containing 50 µg/ml kanamycin. Colonies were then picked from each plate and grown up in 2 ml LB/kanamycin mixture. The tubes shook at 250 rpm in a 37°C incubator overnight. The bacterial cells were then harvested and plasmid DNA was purified using the QIAprep Spin Miniprep Kit (QIAGEN: 27104) and a microcentrifuge. Cloning and transformation success was analyzed by sequencing the plasmid DNA. Based on sequencing results, primers for rapid amplification of cDNA ends (RACE) were designed.

RACE

First-strand cDNA synthesis and RACE were carried out using the SMARTer RACE 5'/3' Kit (Clontech: 634860). The buffer mix for cDNA synthesis contained 4.0 µl 5X First Strand Buffer, 0.5 µl DTT (100 mM), and 1.0 µl DNTPs (20 mM). The 5' cDNA reactions contained 1 µl RNA, 1 µl 5'-CDS Primer A, 9 µl nuclease-free water, and 1 µl SMARTer II A Oligonucleotide. The 3' cDNA reactions contained 1 µl RNA, 1 µl 3'-CDS Primer A, and 10 µl nuclease-free water. The master mix that was added to the denatured RNA was made up of 5.5 µl buffer, 0.5 µl RNase Inhibitor, and 2.0 µl SMARTScribe per reaction. For the RACE reaction, the PCR master mix contained 15.5 µl PCR-Grade Water, 25.0 µl 2X SeqAmp Buffer, and 1.0 µl SeqAmp DNA Polymerase. The RACE sample reactions consisted of 2.5 µl RACE-Ready cDNA, 5 µl 10X Universal Primer Mix, 1 µl 5'/ 3' Gene Specific Primer (10 µM), and 41.5 µl Master Mix. The RACE cycling conditions were as follows: 25 cycles of 94°C for 30 sec, 68°C for 30 sec, and 72°C for 3 min.

RESULTS

PCR and sequencing primers were designed to target Symplectin-like protein sequences in six pelagic mollusc species. The transcriptomes for *P. hoylei*, *D. gigas*, *P. bucephalum*, and *V. infernalis* all returned closely matching hits that could be used for primer design. RNA and DNA were then successfully extracted and purified from tissues from those four species. The transcriptomes for *C. calyx* and *O. deletron* did not return promising alignment results. For that reason, these two species were not involved in future analysis steps.

Pterygioteuthis hoylei

Based on the original screening process, we selected the small squid *Pterygioteuthis hoylei* for further characterization (Silva-Davilla et al. 2013). Initial PCR was run using an annealing temperature range from 51-57°C. This reaction was unsuccessful, returning bands of unanticipated sizes. A second round of PCR was run using different primer combinations and three temperatures (49°C, 51°C, and 53°C). The resultant gel had many promising bands from a variety of primer combinations. Seven bands were cut from the gel and the DNA was sequenced (Figure 2). The successful primer sequences can be seen in Table 2. Of those seven DNA extractions, five returned promising sequence results.

Figure 2: The highlighted bands were all cut from this gel. Bands highlighted in purple returned promising sequencing results, while bands highlighted in yellow did not.

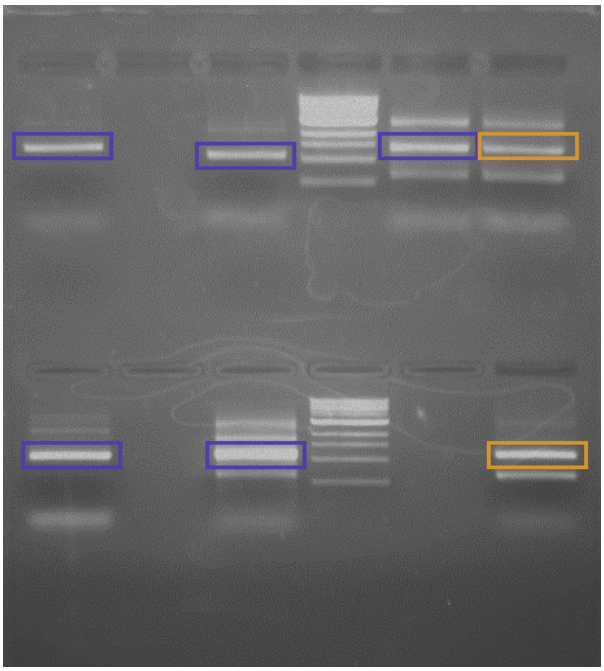
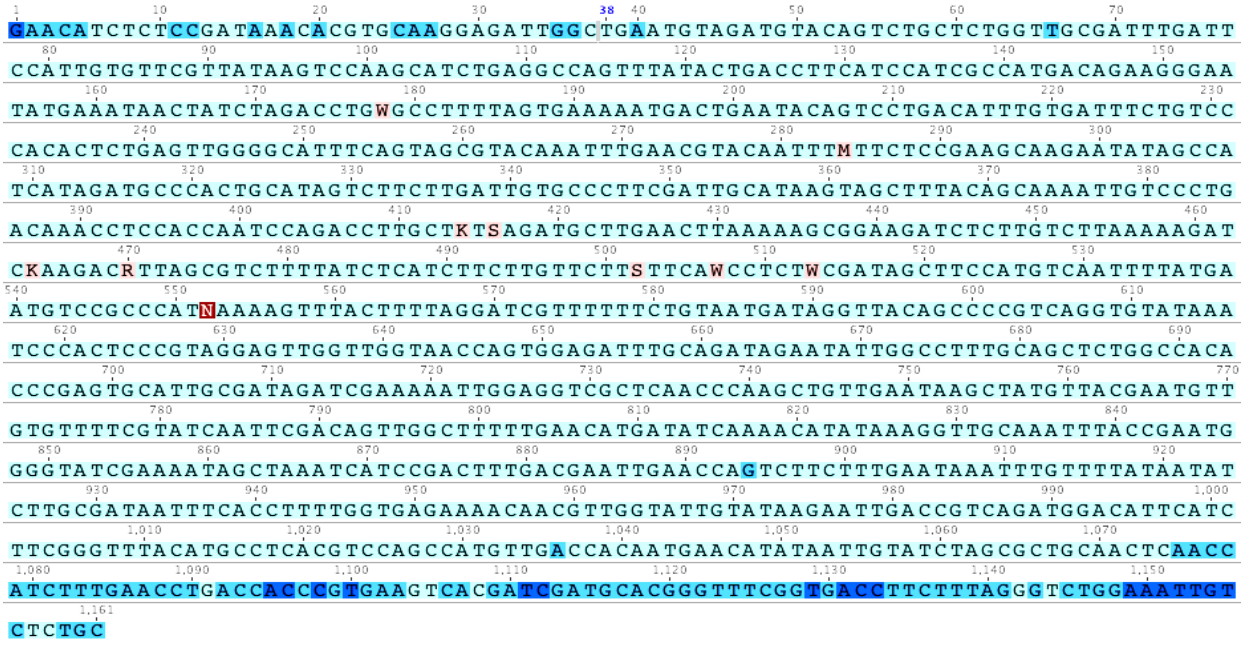


Table 2: This table summarizes primer information for the primers that successfully annealed to and amplified *P. hoylei* DNA. Abbreviations: PtPP = Pterygioteuthis Photoprotein, F = Forward Primer, R = Reverse Primer, IDT = Integrated DNA Technologies, n = nucleotides.

Primer Name	**Sequence**	**GC Content (%)**	**Melting Temperature (°C)**
PtPP 253F - IDT	GGTGTCAAACTGCTGGTTTATC	45.4	54.2
PtPP 461F - IDT	CTGGACGTGAGGCATGTAAA	50.0	54.9
PtPP 1607R - IDT	CCTTCCCAGGCTATCTAGTTTATC	45.8	54.3
PtPP 1328R - IDT	GCCATGACAGAAGGGAATATGA	45.4	54.6
PtPP 1639R - 20n	TTTTGGCAACCTTCCCAGGC	55.0	58.8
PtPP 1639R - 21n	TTTTGGCAACCTTCCCAGGC	52.3	59.9

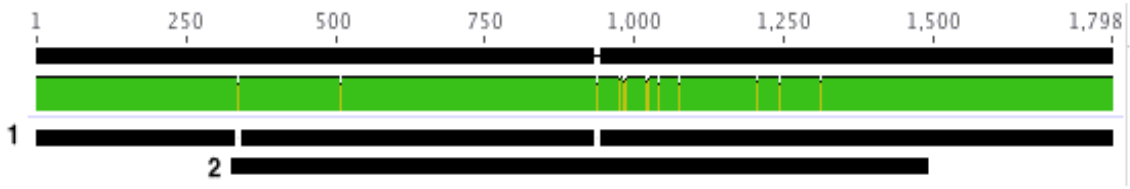
From the sequencing results, a consensus sequence was generated (Figure 3). When this consensus sequence was BLASTed using the NCBI blastx program, the first three hits returned were Symplectin. This suggested that the primers had indeed amplified a Symplectin-like sequence in the DNA as opposed to a random sequence.

Figure 3: A 1,161 base pair consensus sequence was generated from the sequencing results.



Additionally, when the consensus sequence was aligned to the original e^{-147} hit that was used to create the primers, the sequences were almost identical (Figure 4).

Figure 4: Alignment of the original e^{-147} hit (1) to the consensus sequence that was generated from sequencing results (2). Areas of mismatch are denoted in yellow, whereas areas of identical sequence are denoted in green.



The consensus sequence was then BLASTed against all *P. hoylei* transcripts in the transcriptome database. The top ten hits returned from this BLAST were imported into Geneious and aligned with the consensus sequence. From this alignment, it was clear that the extreme 5' and 3' ends of the sequence were still missing, so new primers were designed to amplify each end. From the new batch of primers, DNA was isolated from seven bands on the gel (Figure 5). The successful primer sequences can be seen in Table 3. Sequencing results were returned for six out of the seven DNA extractions. Four different *P. hoylei* PCR products were then cloned and sequenced.

Figure 5: The highlighted bands were all cut from this gel. The bands highlighted in purple returned promising sequencing results, while the band highlighted in yellow did not.

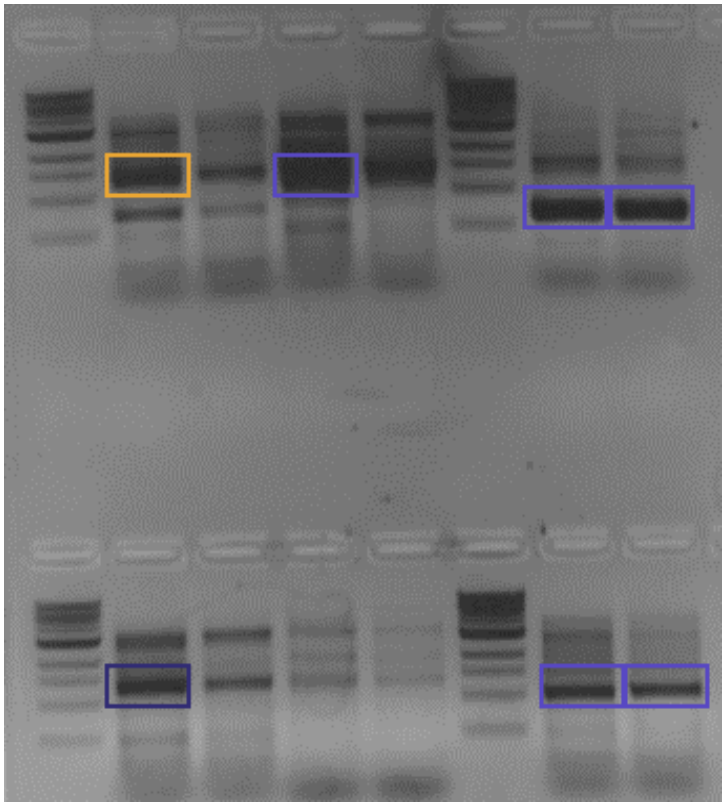


Table 3: This table summarizes primer information for the primers that successfully annealed to and amplified *P. hoylei* DNA. Abbreviations: PtPP = Pterygioteuthis Photoprotein, F = Forward Primer, R = Reverse Primer, IDT = Integrated DNA Technologies, n = nucleotides.

Primer Name	**Sequence**	**GC Content (%)**	**Melting Temperature (°C)**
PtPP 10F - 20n	ATCAACATGTCGGTACATCG	45.0	55.7
PtPP 13F - 21n	AACATGTCGGTACATCGGTTG	47.6	58.7
PtPP 16F - 20n	ATGTCGGTACATCGGTTGTC	50.0	57.5
PtPP 337F - 20n	AGTTGCAGCGCTAGATACAA	45.0	57.3
PtPP 1142R - 20n	CACGTGCAAGGAGATTGGCT	55.0	61.0
PtPP 631R - 20n	AATTTTATGAATGTCCGCCC	40.0	54.3

The clone sequencing results were then assembled to produce nine contigs. Of those nine contigs, only one aligned to Symplectin when BLASTed using the NCBI blastx program. A consensus sequence was generated from the contig, which was originally constructed from eleven reads. Those eleven reads resulted from five different colonies. The forward and reverse sequencing results for each of those five colonies were assembled, resulting in five sequence contigs. Those five sequence contigs were then aligned to produce a strong consensus sequence (Figure 6).

Figure 6: A 1,383 base pair consensus sequence was generated from the clone sequencing results and then translated. A start codon (M) can be seen at position 22 and three stop codons are at positions 449, 453, and 456.

The most 5' region of the consensus sequence, thought to contain the start codon, was BLASTed against all *P. hoylei* assemblies in the transcriptome database. This BLAST search returned several alignments that provided additional 5' and 3' sequence. However, based on analysis of the open reading frame, we believe that anywhere from 1-22 amino acids-worth of sequence is still missing from the 5' end. For that reason, we designed rapid amplification of cDNA ends (RACE) primers that would extend to the 5' end. cDNA was then synthesized from *P. hoylei* RNA and RACE was done on the cDNA. Unfortunately, the cDNA was not amplified by the RACE procedure. For that reason, additional PCR primers were created for the extreme 5' and 3' ends. While these primers did amplify DNA, sequencing suggested that the DNA was not that of a photoprotein.

Phylliroe bucephalum

Initial PCR was run using three different temperatures (43°C, 45°C, and 47°C). The resultant gel showed no bands for any primer combinations. Given the evolutionary distance separating *P. bucephalum* (a pelagic nudibranch) from *S. oualaniensis* (a midwater squid), this result was unsurprising.

Vampyroteuthis infernalis

The first round of PCR looked at all possible PCR primer combinations at three different temperatures (49°C, 51°C, and 53°C). The resultant gel did not have any bands and showed heavy smearing due to a high concentration of genomic DNA. For the next round of PCR, the genomic DNA was diluted and additional primer combinations (utilizing sequencing primers as well as PCR primers) were tested at the same temperatures. When the PCR products were analyzed on an agarose gel, several bands were produced. DNA was extracted from the two most promising bands and then sequenced (Figure 7). The successful primers are listed in Table 4.

Figure 7: The two highlighted bands were cut from the gel and returned sequencing results.

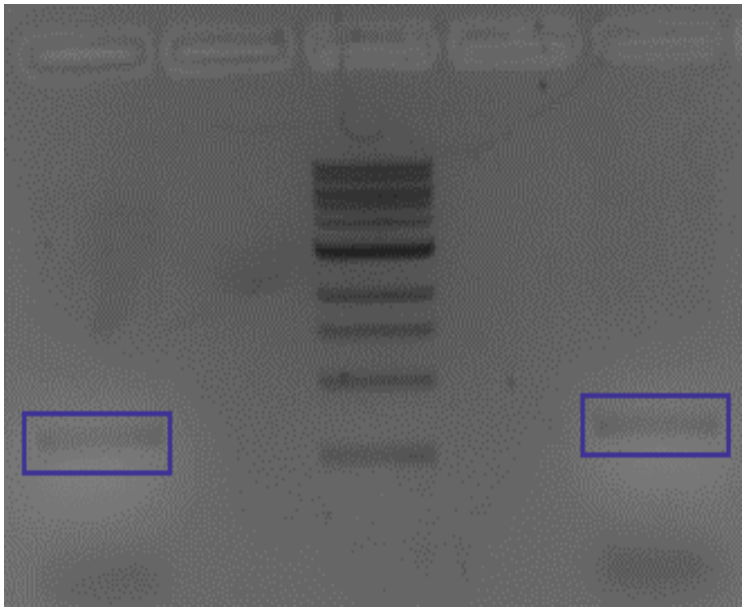


Table 4: This table summarizes primer information for the primers that successfully annealed to and amplified *V. infernalis* DNA. Abbreviations: VtPP = Vampyroteuthis Photoprotein, F = Forward Primer, R = Reverse Primer, IDT = Integrated DNA Technologies.

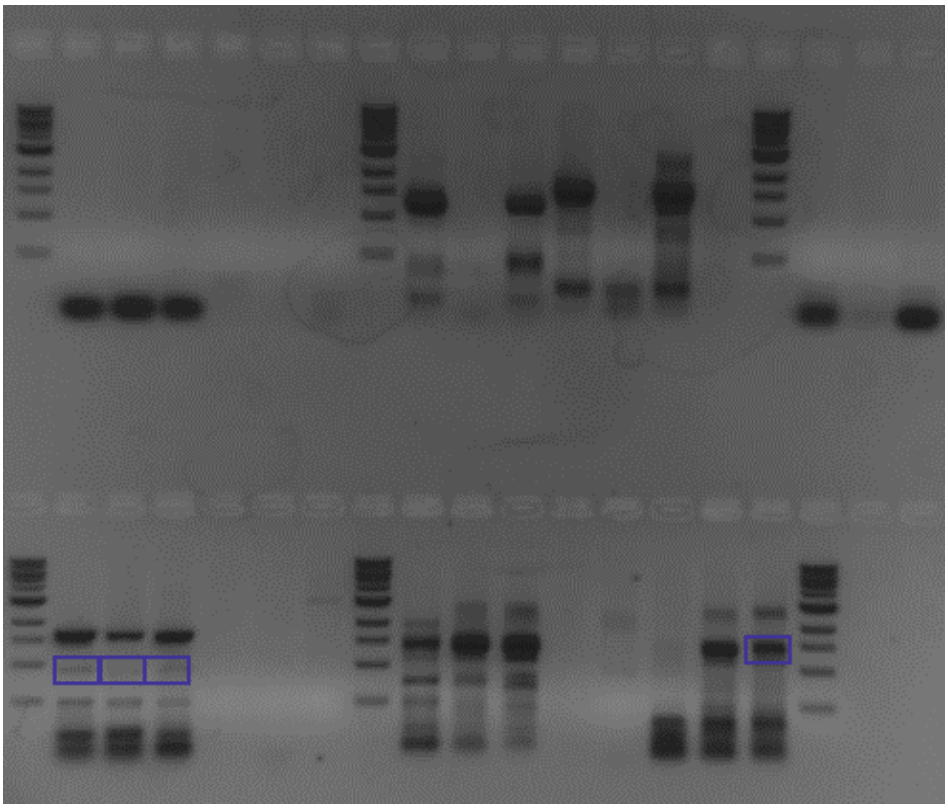
Primer Name	**Sequence**	**GC Content (%)**	**Melting Temperature (°C)**
VtPP 468F - IDT	TACCAGATCCGACCAAGATAGA	45.4	54.4
VtPP 1095R - IDT	ACGTCATCGTCCTTCAGATTAC	45.4	54.3

Based on the sequencing results, a consensus sequence was constructed. However, when the consensus sequence was input into the NCBI blastx program, the top alignment hits were not to Symplectin. Additionally, when the consensus sequence was aligned to the original hit used in primer design, it was clear that a large portion of both 5' and 3' sequence was missing. One PCR product was cloned and sequenced. Sequence assembly produced two contigs, neither of which aligned back to Symplectin when BLASTed. Additionally, assembly of the forward and reverse sequence data for the colonies that contributed to the contigs failed to produce a consensus sequence that aligned to Symplectin.

Dosidicus gigas

Initial PCR using a variety of combinations of forward and reverse primers resulted in a gel that had bands at 1500 bp, 2500 bp, and 3000 bp. The expected fragment size was approximately 1500 bp. The successful forward and reverse primer pair underwent a second round of PCR, this time using a temperature gradient. The gel showed that the optimal annealing temperature for this primer pair was 55°C. The remaining PCR product from the 55°C reaction was loaded onto a gel and all three bands were excised and sequenced, but no viable sequencing results were returned. In a third round of PCR, all remaining primer combinations were tested at a range of temperatures from 49°C to 55°C. The gel showed thirteen promising bands, of which four were eventually gel cut (Figure 8).

Figure 8: The four highlighted bands were cut from the gel and returned sequencing results.



One PCR product was later determined to be a viable candidate for cloning. While cloning did return sequencing results, the consensus sequence did not show significant similarity to anything in the NCBI blastx database.

DISCUSSION

Initially, six species were chosen for analysis in this study: *P. hoylei*, *D. gigas*, *P. bucephalum*, *V. infernalis, C. calyx* and *O. deletron.* Five of these species are deep-sea cephalopods while *P. bucephalum* is a pelagic nudibranch. These six species were chosen because they are all known to produce bioluminescence. However, little is known about the photoproteins/luciferases responsible for bioluminescence in each of these species. Additionally, novel transcriptomic data was available for each species through an in-house transcriptome database. Thus the overall goal of the project was to better characterize the photoprotein/luciferase genetics of these species by comparing transcriptomic data to a known cephalopod photoprotein sequence (Symplectin). In this way, we hoped to contribute to a greater understanding of bioluminescence in pelagic molluscs.

Initial transcriptomic analysis for *P. hoylei*, *D. gigas*, *P. bucephalum*, and *V. infernalis* returned promising alignments to the Symplectin photoprotein sequence. However, when the Symplectin sequence was BLASTed against the transcriptomes for *C. calyx* and *O. deletron,* no strong alignment results were returned. This null result suggests that the photoproteins/luciferases in *C. calyx* and *O. deletron* are genetically distinct from Symplectin. In turn, this implies that *C. calyx* and *O. deletron* have undergone divergence from *S. oualaniensis* in terms of bioluminescent machinery. On face value, this may not initially seem logical, given that both *V. infernalis* and *P. bucephalum* returned promising results but do not share an order (Oegopsida) with *S. oualaniensis*, while *C. calyx* and *O. deletron* do share an order but did not produce alignment results. However, it is generally accepted that bioluminescence has independently evolved at least seven times within the Mollusca phylum alone, offering support for this result (Haddock et al. 2009).

Based on alignments between Symplectin and different transcriptome hits, primers were designed to locate Symplectin-like sequence in the DNA of *P. hoylei*, *D. gigas*, *P. bucephalum*, and *V. infernalis.* However, the primers for *P. bucephalum* failed to amplify DNA. Given the evolutionary distance between *P. bucephalum* and *S. oualaniensis*, this result was unsurprising. It is likely that the photoprotein/luciferase involved in bioluminescence in *P. bucephalum* is genetically and structurally distinct from Symplectin. In contrast, PCR was successful for *P. hoylei*, *D. gigas*, and *V. infernalis.* However, these successful PCRs against gDNA did not turn up any introns in the specimens. This result is interesting and could explain why PCR did not work for *P. bucephalum*. Based on these promising results, cloning and transformation were carried out in the hopes of gathering more complete sequence for each of the three remaining species. The sequencing results for *D. gigas* and *V. infernalis* suggested that PCR had amplified something other than Symplectin in the DNA. Again, this result suggests that

these two species have developed bioluminescence machinery different from that of *S. oualaniensis*.

However, the *P. hoylei* clone sequencing results strongly suggest that a Symplectin-like photoprotein is present in *P. hoylei*. Overall, sequencing has produced 1,387 bases of viable sequence from the *P. hoylei* genomic DNA. When this sequence is translated, a clear open reading frame is immediately apparent. Based on this open reading frame, we are confident that we have succeeded in determining the sequence at the 3' end of the photoprotein transcript, given the presence of three stop codons. This conclusion is supported by transcriptomic comparison as well. However, while the 5' end does have a relatively early start codon, transcriptomic comparison suggests that 1-20 amino acids-worth of sequence are still missing. The real start codon most likely lies somewhere in that upstream region.

CONCLUSIONS/RECOMMENDATIONS

In conclusion, our results suggest that the photoprotein/luciferase responsible for bioluminescence in *P. hoylei* shares strong genetic similarities with Symplectin, a photoprotein derived from *S. oualaniensis*. Our results also suggest that the photoproteins/luciferases involved in bioluminescence in *D. gigas*, *P. bucephalum*, and *V. infernalis* may share some similarities to Symplectin (as evidenced by transcriptome BLAST hits). However, *C. calyx* and *O. deletron* did not return transcriptome BLAST hits and are therefore most likely genetically diverged from Symplectin. Future studies should continue gathering sequence data from *P. hoylei,* particularly at the 5' end. This could be accomplished by doing rapid amplification of cDNA ends (RACE). While our initial attempts at RACE were unsuccessful, there are several factors that could be altered to increase the chances of success. For example, primers should be redesigned to ensure that they bind to the targeted DNA regions. Additionally, the temperature parameters of the reaction could be optimized to get better results. Once the full length sequence of the transcript has been determined, the next logical step would be to insert the transcript into competent *E. coli* cells and attempt to express the photoprotein/luciferase. In turn, this would allow for further characterization of bioluminescence in *P. hoylei*. Overall, future studies of a similar nature will allow a better understanding of light in the deep sea and the organisms that produce it.

ACKNOWLEDGEMENTS

I would like to sincerely thank Steve Haddock for his guidance and support throughout the 2014 internship program. In addition, I would like to thank Lynne Christianson for her tireless assistance and advice. I would also like to thank Warren Francis and April Woods for always sharing their experience and wisdom. In addition, I thank George Matsumoto and Linda Kuhnz for their support throughout the internship program. This work was generously supported by the David and Lucile Packard Foundation and the Monterey Bay Aquarium Research Institute.

REFERENCES

Bjorn, L.O. (2008). The science of life and light. *Springer Netherlands*, **2**: 1-684.

Bush, S.L., Robison, B.H., and R.L. Caldwell (2009). Behaving in the dark: Locomotor, chromatic, postural, and bioluminescent behaviors of the deep-sea squid *Octopoteuthis deletron*. *The Biological Bulletin,* **216**: 7-22.

Clarke, M.R. (1965). Large light organs on the dorsal surfaces of the squids *Ommastrephes pteropus, Symplectoteuthis oualaniensis,* and *Dosidicus gigas*. *Proceedings of the Malacological Society London*, **36**: 319-21.

Fujii, T., Ahn, J., Kuse, M., Mori, H., Matsuda, T., and M. Isobe (2002). A novel photoprotein from oceanic squid (*Symplectoteuthis oualaniensis*) with sequence similarity to mammalian carbon-nitrogen hydrolase domains. *Biochemical and Biophysical Research Communications*, **293**: 874-79.

Haddock, S.H.D., Moline, M.A., and J.F. Case (2010). Bioluminescence in the sea. *Annual Review of Marine Science,* **2**: 443-93.

Herring, P.J. (1987). Systematic distribution of bioluminescence in living organisms. *Journal of Bioluminescence and Chemiluminescence,* **1**: 147-63.

Silva-Davilla, R., Hochberg, F.G., and A.R. Lindgren (2013). Paralarval development, abundance, and distribution of *Pterygioteuthis hoylei* (Cephalopoda: Oegopsida: Pyroteuthidae) in the Gulf of California, Mexico. *Molluscan Research*, **33**(1): 50-65.

Takahasi, H. and M. Isobe (1993). *Symplectoteuthis* bioluminescence (1) Structure and binding form of chromophore in photoprotein of a luminous squid. *Bioorganic and Medicinal Chemistry Letters,* **3**: 2647-52

Teranishi, K. and O. Shimomura (2008). Bioluminescence of the arm light organs of the luminous squid *Watasenia scintillans*. *Biochimica et Biophysica Acta,* **1780**(5): 784-92.

Francis, W.R., Christianson, L.M., Kiko, R., Powers, M.L., Shaner, N.C., and S.H.D. Haddock (2013). A comparison across non-model animals suggests an optimal sequencing depth for *de novo* transcriptome assembly. BMC Genomics, **14**:167.