Luciferin production and luciferase transcription in the bioluminescent copepod *Metridia lucens*

Michael Tessler[1], Jean P. Gaffney[2,3], Jason M. Crawford[4], Eric Trautman[4], Nehaben A. Gujarati[2], Philip Alatalo[5], Vincent A. Pieribone[6] and David F. Gruber[2,3]

[1] Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, NY, USA
[2] Department of Natural Sciences, City University of New York, Bernard M. Baruch College, New York, NY, United States of America
[3] Biology, City University of New York, Graduate School and University Center, New York, NY, United States of America
[4] Department of Chemistry, Yale University, New Haven, CT, United States of America
[5] Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA, United States of America
[6] Cellular and Molecular Physiology, Yale University, New Haven, CT, United States of America

Corresponding authors
Michael Tessler, mtessler@amnh.org
Jean P. Gaffney,
jean.gaffney@baruch.cuny.edu

ABSTRACT

Bioluminescent copepods are often the most abundant marine zooplankton and play critical roles in oceanic food webs. *Metridia* copepods exhibit particularly bright bioluminescence, and the molecular basis of their light production has just recently begun to be explored. Here we add to this body of work by transcriptomically profiling *Metridia lucens*, a common species found in temperate, northern, and southern latitudes. In this previously molecularly-uncharacterized species, we find the typical luciferase paralog gene set found in *Metridia*. More surprisingly, we recover noteworthy putative luciferase sequences that had not been described from *Metridia* species, indicating that bioluminescence produced by these copepods may be more complex than previously known. This includes another copepod luciferase, as well as one from a shrimp. Furthermore, feeding experiments using mass spectrometry and ^{13}C labelled L-tyrosine and L-phenylalanine firmly establish that *M. lucens* produces its own coelenterazine luciferin rather than acquiring it through diet. This coelenterazine synthesis has only been directly confirmed in one other copepod species.

INTRODUCTION

Copepods are small crustaceans (generally ~1–2 mm) that are a primary component of marine zooplankton communities, holding key roles in oceanic food webs (*Takenaka, Yamaguchi & Shigeri, 2017*). These animals act as lower-order consumers, important grazers on microplankton, and as a major food source for many invertebrate and vertebrate predators, such as cephalopods and fishes (*Marine Zooplankton Colloquium 2, 2001*; *Takenaka, Yamaguchi & Shigeri, 2017*). Globally, the bulk of plankton communities (generally 55–95% of the individuals) are often populated by copepods (*Longhurst, 1985*). Within
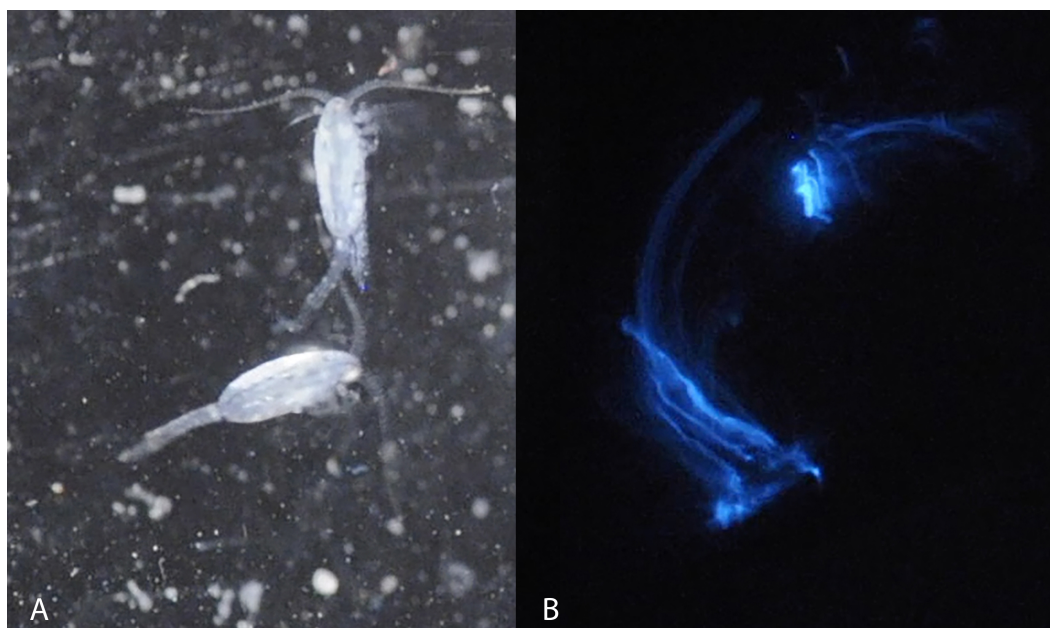
Figure 1 **Light microscopy of *Metridia lucens*.** Under normal light (A) and while excreting biolumi-nesce (B).

the zooplanktonic copepods, the order Calanoida can dominate (e.g., in the Arctic Ocean) in terms of biomass (*Kosobokova & Hirche, 2009*; *Forest et al., 2011*). Given the exceptional biomass and low-level trophic positioning of the calanoids, these tiny drifters themselves act as an important component of aquatic foodwebs (*Takenaka, Yamaguchi & Shigeri, 2017*).

Within the calanoid copepods, bioluminescent species can comprise much of the life (5–59% of the abundance and 10–15% of the biomass) in marine ecosystems (*Takenaka, Yamaguchi & Shigeri, 2017*). Species in the superfamily Augaptiloidea are known for their bioluminescence, and in at least one comparison they produced the highest bioluminescent activity of any calanoids (*Takenaka et al., 2012*). Species in the genus *Metridia* have exceptionally bright bioluminescence (*Takenaka et al., 2012*). Several *Metridia* species have been reviewed regarding the molecular basis of their bioluminescent activity, finding interesting patterns of luciferase gene duplication within the Metridinidae (*Takenaka et al., 2012*). Still, most of the species remain poorly studied or are uncharacterized in this regard.

Blue bioluminescence produced by *Metridia lucens* (Fig. 1 and Supplemental Informa-tion) was first noted when the species was described over 150 years ago (*Boeck, 1864*), yet its bioluminescence still remains uncharacterized molecularly. This moderately sized copepod (2.4–3.0 mm for females and 1.8–2.5 mm for males) is common in temperate to more northerly waters, as well as southerly waters (*David & Conover, 1961*; *Stupnikova et al., 2013*). Both anterior and posterior regions of *M. lucens* bioluminesce, ultimately producing a bright flash that tapers off over a few seconds but can be activated repeatedly (*David & Conover, 1961*). The bioluminescent flashes produced by *M. lucens* can be triggered by disturbance to the water and possibly by the presence of predators (*David & Conover, 1961*).

The two primary photoreceptors found in Crustacea are naupliar eyes and compound eyes. *Metridia lucens* and all copepods use naupliar eyes which are traditionally considered simpler structures that function in phototaxis, rather than in image formation (*Ali, 2013*). However, naupliar eyes are not always simple and can be highly specialized in some groups of copepods (*Porter et al., 2017*). Despite not having image-forming eyes, copepods react strongly when exposed to simulated bioluminescence (*Buskey & Swift, 1985*; *Widder & Johnsen, 2000*). Like several other copepods, *M. lucens* increase their swimming speed and spiral off photophobically in an overall straight path when exposed to such flashes (*Buskey & Swift, 1985*).

In the Gulf of Maine, *M. lucens* densely aggregate, forming bioluminescent "hot spots" that are vertically constrained to a half meter (*Widder et al., 1999*). Given the response of *M. lucens* to physical disturbance, the half meter band where they aggregate acts as a "bioluminescent minefield" (*Widder & Johnsen, 2000*; *Widder, 2002*). It has been hypothesized that when a larger animal disturbs the *M. lucens* aggregation, it will trigger the copepods to illuminate the water and expose this larger animal to even larger predators that may be lurking (*Widder & Johnsen, 2000*; *Widder, 2002*). While the exact mechanism should be tested further, it is evermore apparent that bioluminescence in *M. lucens* is at least in part some form of predator evasion.

In this study we seek to characterize molecular aspects of bioluminescence in the copepod *M. lucens*. To accomplish this, we performed Illumina-based transcriptome sequencing of this species to find RNA sequences of luciferase enzymes and to search for putative luciferin producing sequences. Furthermore, we review this species for biomolecular products relating to the luciferins used as substrate for bioluminescence.

MATERIALS AND METHODS

Sample collection

Metridia lucens samples were collected on January 12, 2016 using the *R/V Tioga* (Woods Hole Oceanographic Institution) on Cape Cod Bay in waters of 30–40 m depth. Their identities were confirmed morphologically and with COI and 18S transcriptomic sequences. A $\frac{3}{4}$-m plankton net with 243 μm mesh netting was deployed and a double-oblique tow was conducted to 2 m off the bottom. Contents of each tow were diluted into ambient seawater in 30-liter barrels or into insulated foam coolers for transport to the laboratory. Within four hours, plankton was transferred to a walk-in cold-room for sorting (at a rate of 50 copepods/hour). Cultured phytoplankton (a mixture of diatoms, flagellates, and dinoflagellates) were fed to the copepod cultures, held in 12-L glass carboys, with diffuse air-stones for aeration at 8 °C, and a photoperiod of 13L/11D. Cultures were changed every three days and checked for spawned eggs. Eggs were removed and cultured separately, as adults are known to be cannibalistic on eggs and nauplii.

Transcriptomic search for bioluminescence related sequences

RNA was extracted from 10 individuals of *M. lucens* using an RNeasy Fibrous Tissue Mini Kit (Qiagen #74704); males and females were mixed in the collection used. Extracted RNA was then used for transcriptome sequencing at the New York University Langone

Medical Center Genomics Technology Center using an Illumina HiSeq 2500 with paired-end 100 bp sequences (using V4 chemistry). The resulting sequences were processed using our prior protocols (*Tessler et al., 2018*), which are briefly summarized here. First, sequences were trimmed for quality control using Trimmomatic (*Bolger, Lohse & Usadel, 2014*) and then assembled using Trinity 2.4 (*Grabherr et al., 2011*; *Haas et al., 2013*) with default parameters. Open reading frames (ORFs) were then predicted from the assembled transcriptome using Transdecoder 3.0 (*Haas et al., 2013*), with a minimum requirement of 5 amino acids due to our interest in potentially short luciferins.

ORFs were then queried against three local databases. The first and primary local database consisted of crustacean luciferase sequences from GenBank, with a focus on copepod luciferases. The second local database consisted of photoprotein sequences used in *Brugler et al. (2018)*. The third local database consisted of potential luciferins and closely related isopenicillin-N-synthase sequences. The putative luciferin sequences in this third database ended in the amino acids ''FYY'', were found to be similar to isopenicillin-N-synthase, and were identified as putative luciferin biosynthesis proteins in prior work (*Francis et al., 2015*). For each query against these local databases we used blastp with e^{-5} set as an e-value minimum; only the top matching hit was retained for further review. To insure that no other proteins were better matches to ORFs putatively matching a sequence in our local database, we used these ORFs as queries against the Swissprot/Uniprot database using blastp. The Swissprot/Uniprot database was used here rather than other databases because it is well-curated. When bitscores from the local database queries were equal or higher than the results from Swissprot/Uniprot, the sequence were considered a putative match to the protein of interest (e.g., candidate luciferases). Bitscores were used, as they avoid issues of differing database size (here our local database vs. Swissprot/Uniprot) inherent in e-value cutoffs.

Phylogenetics of *Metridia*-related Luciferases and COI

In order to produce a *Metridia*-related luciferase gene tree, luciferase sequences utilized for a prior phylogenetic reconstruction (*Takenaka et al., 2012*) were combined with our ORF sequences that best matched *Metridia* sequences from that study; specifically, the single longest ORF per Trinity read cluster. Alignments of these sequences were then produced using MUSCLE v3.8.31 (*Edgar, 2004*). A model of amino acid replacement was selected (WAG + G) for the resulting alignment using ProtTest 3.4.2 (*Darriba et al., 2011*). A phylogenetic reconstruction with 1,000 bootstrap replicates for support was then produced using the aligned amino acid sequences and selected model in RAxML 8.2.10 (*Stamatakis, 2014*) in the CIPRES Science Gateway (*Edgar, 2004*; *Miller, Pfeiffer & Schwartz, 2010*). Outgroup taxa (*Heterorhabdus tanneri*, *Heterostylites major*, and *Lucicutia ovaliformis*) for the resulting luciferase gene tree were selected based on a previous luciferase phylogenetic reconstruction of bioluminescent copepods (*Takenaka et al., 2012*).

As the taxonomic status of *M. lucens* has been debated in the literature, we generated a phylogenetic matrix of the COI barcoding locus for *Metridia* (Supplemental Information). Sequences downloaded from GenBank had their primers trimmed when necessary. Phylogenetic reconstruction follows that of the luciferases, except nucleotides were used

Table 1 Crustacean luciferases that match sequences from the transcriptome of *Metridia lucens*.

Taxon	GenBank accession	Percent identity	Matching length	*E*-value	Bitscore
Heterorhabdus tanneri	BAL63040	36.8	38	5.4E−06	33.5
Metridia okhotensis	BAL63033	97.2	144	2.3E−106	293
Metridia pacifica	BAD93334	96.9	163	10.0E−117	320
Metridia pacifica	BAY00650	100	96	1.4E−69	198
Metridia pacifica	BAY00651	100	92	2.0E−66	190
Metridia pacifica	BAG48249	78.5	65	3.6E−31	99.4
Metridia pacifica	BAD93333	97.6	41	1.5E−25	84
Metridia pacifica	BAY00656	94.7	19	4.4E−11	44.7
Oplophorus gracilirostris	Q9GV46	25.5	376	1.3E−23	92
Pleuromamma sp.	AAG54096	62.1	124	6.6E−57	167

and jModelTest 2.1.4 (*Darriba et al., 2012*) determined a model (GTR + I + Γ). The outgroups were furthermore based on the 18S phylogeny of bioluminescent copepods (*Takenaka et al., 2012*).

Isotope experiment

^{13}C labelled L-tyrosine and L-phenylalanine (Cambridge Isotopes CLM-2263 and CLM-2250) were added to seawater in a final concentration of 10 mg/L together with live *M. lucens* specimens. The copepods were not fed once they were transferred to seawater containing isotopes. Copepods remained in the labelled seawater for 4 days. On Day 4, nine individuals were collected from the control, five from the phenylalanine culture, and one from the tyrosine culture. The whole organisms were frozen immediately upon collection. *Metridia lucens* were lyophilized as received. The remaining solids were extracted with MeOH +1 M HCl. Extracts were analyzed by LC-MS using an Agilent iFunnel 6550 Quadrupole-Time-of-Flight Mass Spectrometer (Q-TOF).

RESULTS

Transcriptomic search for bioluminescence related sequences

Raw transcriptomic reads are deposited in the Short Read Archive under accession SRX3899629. The assembled transcriptomic sequences had 308,066 contigs, that resulted in 237,343 ORFs. Of the luciferase sequences in our local database, 10 had matching ORFs that were retained after reciprocal BLAST searches against Swissprot/Uniprot (Table 1 and Supplemental Information). Of these, nine were matches to other species in the copepod; specifically, *Heterorhabdus tanneri*, *Metridia okhotensis*, *Metridia pacifica*, and *Pleuromamma* sp. Another sequence matched a bioluminescent shrimp (*Oplophorus gracilirostris*).

None of the sequences matching putative luciferins were retained after reciprocal BLAST searches against Swissprot/Uniprot and none ended in "FYY" amino acid sequences, as has been suggested to be of importance (*Oba et al., 2009*; *Francis et al., 2015*). Instead these sequences generally better matched other oxidoreductases. Similarly, no photoproteins were uncovered; these better matched calmodulin or, less frequently, troponin.
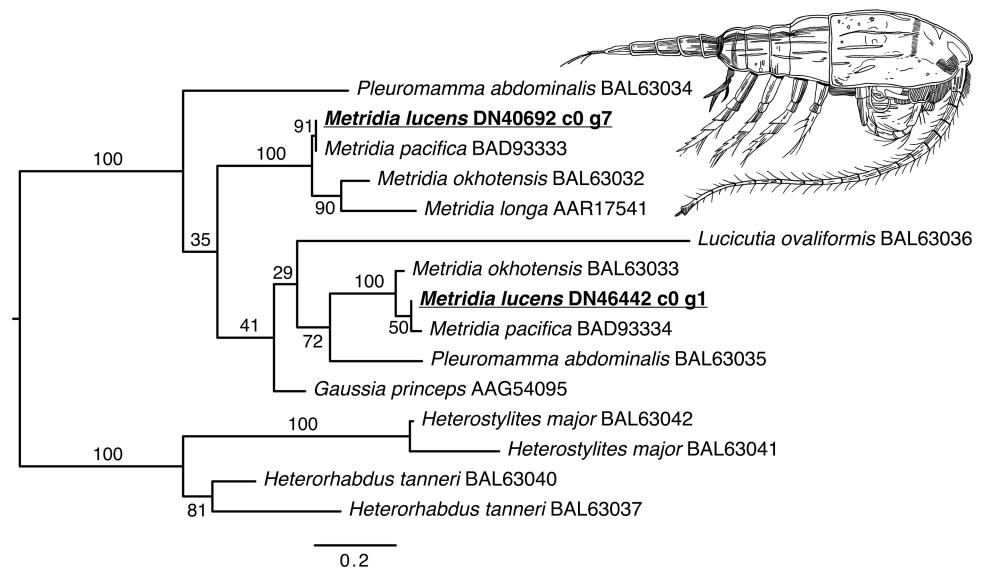
Figure 2 Maximum likelihood phylogenetic reconstruction of luciferases from *Metridia* species and related copepods. Support values are summaries of 1,000 bootstrap replicates. Image credit for copepod drawing: Olivia A. Gresham.

Phylogenetics of *Metridia*-related Luciferases and COI

The phylogenetic reconstruction of the *Metridia*-related luciferases resulted in two principal ingroup clades, each with a paralogous pair of sequence from our ORFs (*M. lucens*), *M. okhotensis*, and *M. pacifica* (Fig. 2). One of these clades also contained single sequences for *M. longa* and *Gaussia princeps*, as well as one of the *Pleuromamma abdominalis* sequences. The *Metridia* sequences in each of these clades were well-supported as being monophyletic (bootstraps = 100%). Many other parts of the tree were poorly supported.

The phylogenetic reconstruction of the COI locus for *Metridia* species (Fig. 3 and Supplemental Information) showed clear differentiation between described species and high genetic variation within *M. lucens* (<10%). We would also like to note that a sequence labeled as *Metridia gerlachei* (HM045328) is clearly a misidentified *M. lucens* individual, and that the sequence for *Metridia venusta* nested within the *Pleuromamma* clade (either this species is misidentified or the taxonomy is in need of updating).

Isotope experiment

Coelenterazine was not detected in any of the samples, likely due to the age of the copepods. However, the degradation product of coelenterazine, coelenteramine was detected (Fig. 4). Phenylalanine and tyrosine labels were observed in coelenteramine from the copepods, supporting that the copepods synthesize coelenterazine.
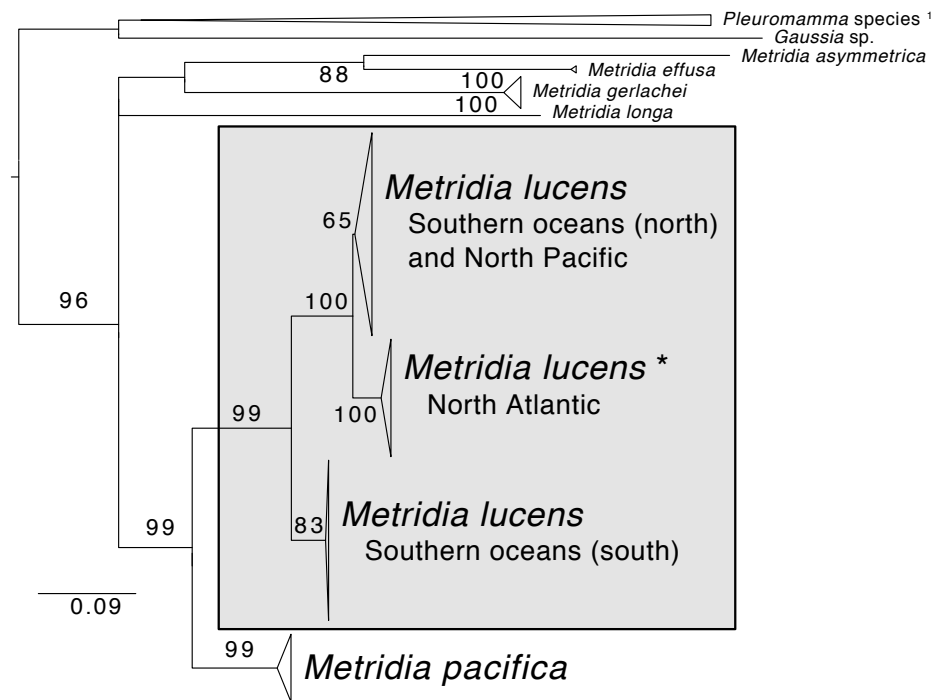
Figure 3 **Maximum likelihood phylogeny of COI of *Metridia* copepods.** Support values are summaries of 1,000 bootstrap replicates. * The North Atlantic *Metridia lucens* clade contains the COI sequence from our *M. lucens* transcriptome. [1]In addition to the *Pleuromamma* species, a sequence of *Metridia venusta* was also found in this clade.

DISCUSSION

Like other bioluminescent *Metridia* species (*Takenaka et al., 2012*), *M. lucens* appears to transcribe luciferase enzymes that oxidize luciferins in order to produce light, and it does not appear to rely on photoproteins for this task. More surprisingly, *M. lucens* transcribes putative luciferases that have not been noted for other *Metridia* copepods. Furthermore, *M. lucens* is one of the few organisms now documented to produce its own coelenterazine luciferin, extending findings on the closely related species *M. pacifica* (*Oba et al., 2009*). Below we detail these findings, resulting in the first molecular characterization of bioluminescence in *M. lucens*.

Luciferases

In our transcriptomic data, several sequences matched luciferases from a variety of crustaceans. Most interestingly, we find a luciferase from the deep-sea shrimp *Oplophorus gracilirostris*. This is notable, as no such sequence has previously been recorded from any species of *Metridia* or other copepod as far as we are aware.

We also recovered a number of sequences matching the more well-reviewed Metridinidae luciferases, corroborating prior work (*Markova et al., 2004*; *Takenaka et al., 2008*; *Takenaka et al., 2012*). A Metridinidae luciferase was first cloned from *Gaussia princeps* (*Verhaegent & Christopoulos, 2002*). Luciferases in this family have more recently been found and
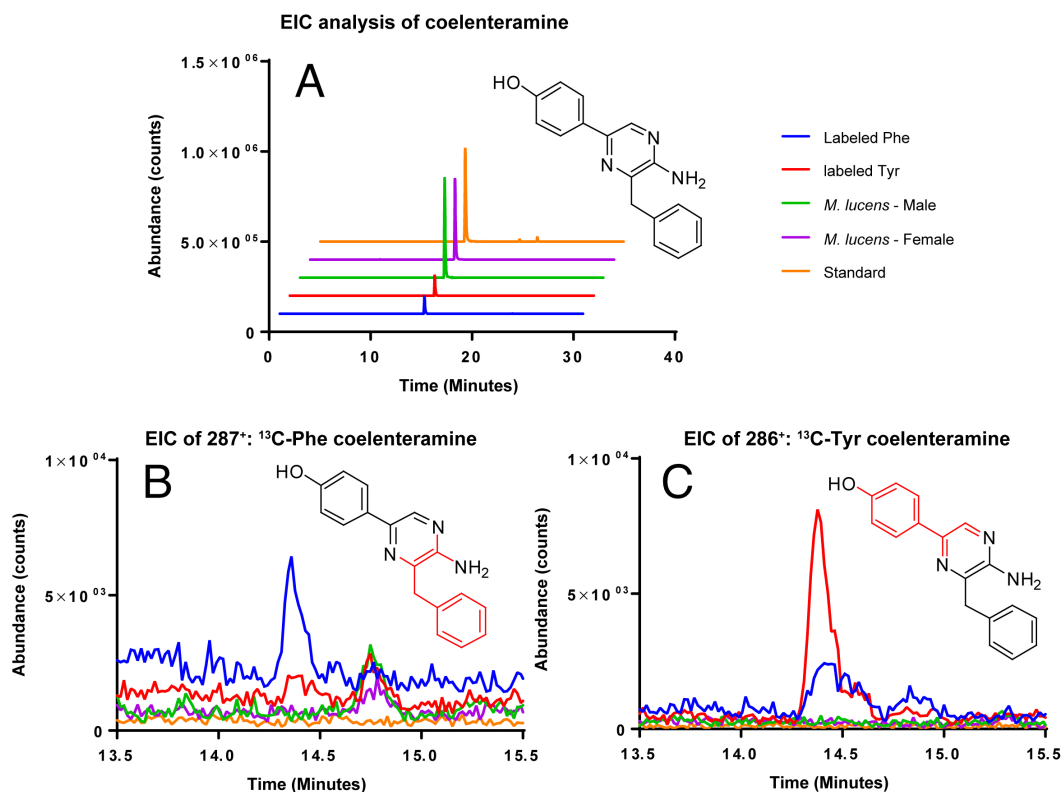
Figure 4 **EIC analysis of the labeled coelenteramines.** (A) Extracted ion count chromatograms (EICs, extraction window ± 10 ppm) for natural ^{12}C-coelenteramine (M + H) under ^{13}C-labeling and non-labeling conditions. (B/C) EICs for the ^{13}C$_9$-labeled (B) and the ^{13}C$_8$-labeled (C) coelenteramine ions centered at m/z 287.1590 and m/z 286.1556 are shown. Data support that the carbons depicted in red are isotopically labeled at low abundance. The peak at 14.4 min has the same retention time as the coelenteramine standard. Observed for ^{13}C$_8$-coelenteramine: 286.1561, 1.7 ppm error. Observed for ^{13}C$_9$-coelenteramine: 287.1598, 2.8 ppm error.

compared for several *Metridia* species (*Takenaka et al., 2012*; *Takenaka et al., 2013*), and have been shown to usually consist of two paralogous pairs of sequences per species. Indeed our *Metridia*-related luciferases fall out into these two clades.

Aside from the typical *Metridia* luciferases, we recovered a match from *M. lucens* to a *Pleuromamma* (Metridinidae) luciferase that had not been uncovered before from a *Metridia* species. While *Pleuromamma abdominalis* does also have the typical luciferase studied in *Metridia* (*Takenaka et al., 2012*), this second luciferase comes from an unidentified member of the genus and has documented bioluminescent activity (*Bryan & Szent-Gyorgyi, 2001*). Our finding of this luciferase in *M. lucens* expands the number of putative luciferases transcribed by the genus *Metridia*. To assess functional similarity, our sequence was modeled using Phyre2 (*Kelley et al., 2015*) and SWISS-MODEL (*Waterhouse et al., 2018*) and compared to the model generated for this *Pleuromamma* luciferase. The Pyre2 models used the same template and showed a confidence level of approximately 33% indicating some similarity, but notable difference. The SWISS-MODEL used different

templates. It is worth noting that *Pleuromamma* is one of the genera most closely related to *Metridia* (*Takenaka et al., 2012*); however, this second luciferase from *Pleuromamma* had not been previously searched for in *Metridia* as far as we are aware.

Luciferins

The majority of bioluminescent organisms react with the luciferin coelenterazine to produce light, including a variety of phyla and close to 100 genera (*Thomson, Herring & Campbell, 1997*; *Haddock, Moline & Case, 2010*; *Markova & Vysotski, 2015*). While many animals acquire the luciferin coelenterazine through diet (*Haddock, Rivers & Robison, 2001*), our mass spectrometry confirms that *Metridia* species appear to be some of the few known animals to produce coelenterazine (*Buskey & Stearns, 1991*; *Oba et al., 2009*). In prior work it was established that *M. pacifica* uses an L-phenylalanine and two L-tyrosine molecules to produce said coelenterazine, possibly with: "autocyclization of tri-peptides such as H_2N-Phe-Tyr-Tyr-COOH" (*Oba et al., 2009*). This study confirms that this second species of *Metridia*, *M. lucens* (or a symbiont it houses), also synthesizes coelenterazine *de novo*.

The promise of coelenterazine being a cyclized "Phe-Tyr-Tyr" ("FYY") led researchers to search for these amino acids at the end of peptides in ctenophores, another animal group believed to produce their own coelenterazine (*Francis et al., 2015*). This resulted in some exciting findings, with C-terminus "FYY" proteins that best matched oxidoreductases: isopenicillin-N-synthases or oxygenases. Furthermore, these "FYY" proteins were found across all bioluminescent species, but were absent from non-bioluminescent ones. Still, while these ctenophore putative oxidoreductases are promising, they have not been expressed or knocked out to confirm function.

None of our ORFs better matched the C-terminus "FYY" oxidoreductases as compared to Swissprot/Uniprot. Furthermore, while we recovered three (Supplemental Information) mature transcripts (here considered to have a start and stop codon) with C-terminus "FYY" sequences, these did not produce any clear matches in BLAST searches (i.e., no well-described matches had an *e*-value below our cutoff). Furthermore, to compare structure, we modeled these ORFs with SWISS-MODEL (*Waterhouse et al., 2018*). This generated dissimilar models, further supporting that these transcript do not share structural similarities with other proteins. However, it is only the most plausible hypothesis thus far put forward that C-terminus "FYY" is required to produce coelenterazine; these amino acids could be produced by other routes. Regardless, there is currently no clear molecular path indicating how *M. lucens* might produce coelenterazine, just that it does produce it.

A taxonomic note

Whether *M. lucens* and *M. pacifica* are one widely distributed species or are indeed two distinct species has been debated in the taxonomic literature (*Bucklin, Frost & Kocher, 1995*). Yet, based on recent molecular studies, the two species appear to be genetically separated by 12.6–14.4% for the COI barcoding locus (*Blanco-Bercial et al., 2014*). Furthermore, there is notable (~8%) genetic differentiation even for lineages within *M. lucens*, suggesting that this taxon alone is composed of several species (*Stupnikova*

et al., 2013; *Blanco-Bercial et al., 2014*). To contextualize this, studies generally consider crustaceans with ~3% genetic divergence in COI to indicate that lineages are separate species (*Stupnikova et al., 2013*). Adding to this evidence, we combined COI sequences of *Metridia* from GenBank and found highly similar results, with *M. lucens* having several distinct clades within it and being genetically distant from *M. pacifica* (10.1–14.6% for COI distance). Our *M. lucens* sample clearly nest within the North Atlantic clade. Future studies should strive to review the lineages within *M. lucens* and determine if morphological traits can be found for describing them as new species.

CONCLUSIONS

Based on our findings, the molecular mechanisms of *M. lucens* bioluminescence appear to be more complex than reported (*Takenaka et al., 2012*) for other species in the genus *Metridia*. Most likely this is not unique to *M. lucens*, but instead *Metridia* and even other genera in the family Metridinidae may have the additional putative luciferases that we recovered for *M. lucens*. Future work should explore the interaction of these bioluminescent loci across Metridiniadae.

ACKNOWLEDGEMENTS

We thank the crew of the *R/V Tioga* (operated by the Woods Hole Oceanographic Institution) for collecting the copepod samples used in this study, Olivia Gresham for providing the line drawing of *Metridia lucens*, and Aida Verdes for assisting with laboratory work.

Funding

This work was supported by DARPA-NESD Grant #N66001-17-C-4012 to Vincent A. Pieribone, David F. Gruber, Jean P. Gaffney, Jason M. Crawford. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
DARPA-NESD Grant: #N66001-17-C-4012.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Michael Tessler conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.

- Jean P. Gaffney conceived and designed the experiments, performed the experiments, contributed reagents/materials/analysis tools, approved the final draft.
- Jason M. Crawford and Vincent A. Pieribone conceived and designed the experiments, performed the experiments, contributed reagents/materials/analysis tools, prepared figures and/or tables, approved the final draft.
- Eric Trautman conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, approved the final draft.
- Nehaben A. Gujarati and Philip Alatalo conceived and designed the experiments, performed the experiments, approved the final draft.
- David F Gruber conceived and designed the experiments, performed the experiments, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.

Data Availability

Raw reads from the *Metridia lucens* transcriptome are deposited in the Short Read Archive under accession SRX3899629 (BioProject PRJNA449123).

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.5506#supplemental-information.

REFERENCES

Ali MA. 2013. *Photoreception and vision in invertebrates.* New York: Plenum Press.

Blanco-Bercial L, Cornils A, Copley N, Bucklin A. 2014. DNA barcoding of marine copepods: assessment of analytical approaches to species identification. *PLOS Currents Tree of Life* **6** DOI 10.1371/currents.tol.cdf8b74881f87e3b01d56b43791626d2.

Boeck A. 1864. Oversigt over de ved Norges Kyster iagttagne Copepoder henhorende til Calanidernes, Cyclopidernes og Harpactidernes Familier af. *Forhandlinger Videnskabs-Selskabet I Cristiana* **7**:226–282.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114–2120 DOI 10.1093/bioinformatics/btu170.

Brugler MR, Aguado MT, Tessler M, Siddall ME. 2018. The transcriptome of the Bermuda fireworm *Odontosyllis enopla* (Annelida: Syllidae): a unique luciferase gene family and putative epitoky-related genes. *PLOS ONE* **13**:e0200944 DOI 10.1371/journal.pone.0200944.

Bryan BJ, Szent-Gyorgyi C. 2001. Luciferases, fluorescent proteins, nucleic acids encoding the luciferases and fluorescent proteins and the use thereof in diagnostics, high throughput screening and novelty items. U.S. Patent 6,232,107.

Bucklin A, Frost BW, Kocher TD. 1995. Molecular systematics of six *Calanus* and three *Metridia* species (Calanoida: Copepoda). *Marine Biology* **121**:655–664 DOI 10.1007/BF00349301.

Buskey EJ, Stearns DE. 1991. The effects of starvation on bioluminescence potential and egg release of the copepod *Metridia* longa. *Journal of Plankton Research* **13**:885–893 DOI 10.1093/plankt/13.4.885.

Buskey EJ, Swift E. 1985. Behavioral responses of oceanic zooplankton to simulated bioluminescence. *The Biological Bulletin* **168**:263–275 DOI 10.2307/1541239.

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**:1164–1165 DOI 10.1093/bioinformatics/btr088.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9(8)**:772 DOI 10.1038/nmeth.2109.

David CN, Conover RJ. 1961. Preliminary investigation on the physiology and ecology of luminescence in the copepod, *Metridia lucens*. *The Biological Bulletin* **121**:92–107 DOI 10.2307/1539462.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**:1792–1797 DOI 10.1093/nar/gkh340.

Forest A, Galindo V, Darnis G, Pineault S, Lalande C, Tremblay J-E, Fortier L. 2011. Carbon biomass, elemental ratios (C:N) and stable isotopic composition (δ_{13}C, δ_{15}N) of dominant calanoid copepods during the winter-to-summer transition in the Amundsen Gulf (Arctic Ocean). *Journal of Plankton Research* **33**:547–547 DOI 10.1093/plankt/fbr003.

Francis WR, Shaner NC, Christianson LM, Powers ML, Haddock SHD. 2015. Occurrence of isopenicillin-N-synthase homologs in bioluminescent ctenophores and implications for coelenterazine biosynthesis. *PLOS ONE* **10(6)**:e0128742 DOI 10.1371/journal.pone.0128742.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**:644–652 DOI 10.1038/nbt.1883.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* **8**:1494–1512 DOI 10.1038/nprot.2013.084.

Haddock SHD, Moline MA, Case JF. 2010. Bioluminescence in the Sea. *Annual Review of Marine Science* **2**:443–493 DOI 10.1146/annurev-marine-120308-081028.

Haddock SH, Rivers TJ, Robison BH. 2001. Can coelenterates make coelenterazine? Dietary requirement for luciferin in cnidarian bioluminescence. *Proceedings of the National Academy of Sciences of the United States of America* **98**:11148–11151 DOI 10.1073/pnas.201329798.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**:845–858 DOI 10.1038/nprot.2015.053.

Kosobokova K, Hirche H-J. 2009. Biomass of zooplankton in the eastern Arctic Ocean—A base line study. *Progress in Oceanography* **82**:265–280 DOI 10.1016/j.pocean.2009.07.006.

Longhurst AR. 1985. The structure and evolution of plankton communities. *Progress in Oceanography* **15**:1–35 DOI 10.1016/0079-6611(85)90036-9.

Marine Zooplankton Colloquium 2. 2001. Future marine zooplankton research-a perspective. *Marine Ecology Progress Series* **222**:297–308 DOI 10.3354/meps222297.

Markova SV, Golz S, Frank LA, Kalthof B, Vysotski ES. 2004. Cloning and expression of cDNA for a luciferase from the marine copepod *Metridia longa*. A novel secreted bioluminescent reporter enzyme. *The Journal of Biological Chemistry* **279**:3212–3217 DOI 10.1074/jbc.M309639200.

Markova SV, Vysotski ES. 2015. Coelenterazine-dependent luciferases. *Biochemistry. Biokhimiia* **80**:714–732 DOI 10.1134/S0006297915060073.

Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *2010 Gateway Computing Environments Workshop (GCE)*.

Oba Y, Kato S-I, Ojika M, Inouye S. 2009. Biosynthesis of coelenterazine in the deep-sea copepod, *Metridia pacifica. Biochemical and Biophysical Research Communications* **390**:684–688 DOI 10.1016/j.bbrc.2009.10.028.

Porter ML, Steck M, Roncalli V, Lenz PH. 2017. Molecular characterization of copepod photoreception. *The Biological Bulletin* **233**:96–110 DOI 10.1086/694564.

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

Stupnikova AN, Molodtsova TN, Mugue NS, Neretina TV. 2013. Genetic variability of the *Metridia lucens* complex (Copepoda) in the Southern Ocean. *Journal of Marine Systems* **128**:175–184 DOI 10.1016/j.jmarsys.2013.04.016.

Takenaka Y, Masuda H, Yamaguchi A, Nishikawa S, Shigeri Y, Yoshida Y, Mizuno H. 2008. Two forms of secreted and thermostable luciferases from the marine copepod crustacean, *Metridia pacifica. Gene* **425**:28–35 DOI 10.1016/j.gene.2008.07.041.

Takenaka Y, Noda-Ogura A, Imanishi T, Yamaguchi A, Gojobori T, Shigeri Y. 2013. Computational analysis and functional expression of ancestral copepod luciferase. *Gene* **528**:201–205 DOI 10.1016/j.gene.2013.07.011.

Takenaka Y, Yamaguchi A, Shigeri Y. 2017. A light in the dark: ecology, evolution and molecular basis of copepod bioluminescence. *Journal of Plankton Research* **39**:369–378 DOI 10.1093/plankt/fbx016.

Takenaka Y, Yamaguchi A, Tsuruoka N, Torimura M, Gojobori T, Shigeri Y. 2012. Evolution of bioluminescence in marine planktonic copepods. *Molecular Biology and Evolution* **29**:1669–1681 DOI 10.1093/molbev/mss009.

Tessler M, Marancik D, Champagne D, Dove A, Camus A, Siddall ME, Kvist S. 2018. Marine leech anticoagulant diversity and evolution. *The Journal of Parasitology* **104**:210–220 DOI 10.1645/17-64.

Thomson CM, Herring PJ, Campbell AK. 1997. The widespread occurrence and tissue distribution of the imidazolopyrazine luciferins. *Journal of Bioluminescence and Chemiluminescence* **12**:87–91 DOI 10.1002/(SICI)1099-1271(199703/04)12:2<87::AID-BIO438>3.0.CO;2-8.

Verhaegent M, Christopoulos TK. 2002. Recombinant Gaussia luciferase. Overexpression, purification, and analytical application of a bioluminescent reporter for DNA hybridization. *Analytical Chemistry* **74**:4378–4385 DOI 10.1021/ac025742k.

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* **46**:W296–W303 DOI 10.1093/nar/gky427.

Widder E. 2002. Bioluminescence and the pelagic visual environment. *Marine and Freshwater Behaviour and Physiology* **35**:1–26 DOI 10.1080/10236240290025581.

Widder EA, Johnsen S. 2000. 3D spatial point patterns of bioluminescent plankton: a map of the minefield. *Journal of Plankton Research* **22**:409–420 DOI 10.1093/plankt/22.3.409.

Widder EA, Johnsen S, Bernstein SA, Case JF, Neilson DJ. 1999. Thin layers of bioluminescent copepods found at density discontinuities in the water column. *Marine Biology* **134**:429–437 DOI 10.1007/s002270050559.