Similar enzymatic functions in distinct bioluminescence systems: Evolutionary recruitment of sulfotransferases in ostracod light organs

Emily S Lau[1], Jessica A Goodheart[1,2], Nolan T Anderson[3], Vannie L Liu[3], Arnab Mukherjee[3,4,5,6*], Todd H Oakley[1*]

[1] Department of Ecology, Evolution, and Marine Biology, University of California Santa Barbara, Santa Barbara, CA 93106

[2] Division of Invertebrate Zoology, American Museum of Natural History, New York, NY 10025

[3] Department of Chemical Engineering, University of California Santa Barbara, Santa Barbara, CA 93106

[4] Department of Biological Engineering, University of California Santa Barbara, Santa Barbara, CA 93106

[5] Department of Chemistry, University of California Santa Barbara, Santa Barbara, CA 93106

[6] Neuroscience Research Institute, University of California Santa Barbara, Santa Barbara, CA 93106

[*] **Corresponding authors:** Arnab Mukherjee (arnabm@ucsb.edu); Todd Oakley (oakley@lifesci.ucsb.edu)

Abstract

Genes from ancient families are sometimes involved in the convergent evolutionary origins of similar traits, even across vast phylogenetic distances. Sulfotransferases are an ancient family of enzymes that transfer sulfate from a donor to a wide variety of substrates, including probable roles in some bioluminescence systems. Here we demonstrate multiple sulfotransferases, highly expressed in light organs of the bioluminescent ostracod *Vargula tsujii*, transfer sulfate in vivo to the luciferin substrate, vargulin. We find luciferin sulfotransferases of ostracods are not orthologous to known luciferin sulfotransferases of fireflies or sea pansies; animals with distinct and convergently evolved bioluminescence systems compared to ostracods. Therefore, distantly related sulfotransferases were independently recruited at least three times, leading to parallel evolution of luciferin metabolism in three highly diverged organisms. Re-use of homologous genes is surprising in these bioluminescence systems because the other components, including luciferins and luciferases, are completely distinct. Whether convergently evolved traits incorporate ancient genes with similar functions or instead use distinct, often newer, genes may be constrained by how many genetic solutions exist for a particular function. When fewer solutions exist, as in genetic sulfation of small molecules, evolution may be more constrained to use the same genes time and again.

Keywords: convergent evolution, parallel evolution, complex traits, gene expression, bioluminescence, sulfotransferase

Introduction

The convergent evolutionary origins of similar traits sometimes employ existing genetic elements that originated much earlier. This pattern of parallel evolution is often called “deep homology”, especially when similar but convergently evolved traits express shared transcription factors. For example, limbs in some distantly related phyla express the transcription factor *distalless* [1]. Convergently evolved traits also may recruit ancient genes with shared enzymatic functions. For example, convergently evolved instances of biomineralization use α-carbonic anhydrases, an ancient gene family found across animals [2] that facilitates conversion of carbon dioxide to bicarbonate, crucial in regulating pH and mineral deposition. Similarly, convergently evolved cases of photosymbiosis involve vacuolar H+-ATPase (*VHA*), another ancient gene family [3,4], which plays a vital role to acidify intracellular compartments, essential for nutrient transport and cellular metabolism. Characterizing new examples where homology differs across levels of biological organization will build the knowledge necessary to answer fundamental evolutionary questions, such as what types of traits, genes, or evolutionary histories are more likely to lead to the use of ancient elements in novel systems [e.g. 5].

We hypothesize sulfotransferases, a family of enzymes shared across vast phylogenetic distances, are used in convergently evolved bioluminescence systems of distantly related species. Fireflies, sea pansies, and ostracods convergently evolved bioluminescence systems and produce light using structurally diverse organs or tissues [6–8], different luciferin substrates [9], and non-homologous, autogenic, luciferase enzymes [10,11] (figure S1). Despite these vast differences, these taxa use the same biochemical mechanism, sulfation — the transfer of sulfate (SO3-) from the donor 3′-phosphoadenosine-5′-phosphosulfate (PAPS) to a substrate [12] — to modulate the chemical state of luciferins [13–16] (figure S2). Sulfation is catalyzed by sulfotransferases in fireflies [15] and sea pansies [16], but less is known about ostracod sulfotransferases. Although previously published results do show crude protein extracts from the ostracod *Vargula hilgendorfii* can reversibly sulfate luciferin, thereby suggesting sulfation as a mechanism for luciferin metabolism in the organism [17], the genetic sequences of luciferin sulfotransferases in ostracods have yet to be identified and functionally characterized.

Here we provide evidence of independent recruitment of paralogous genes from a single, ancient gene family — sulfotransferases — leading to parallel evolution of luciferin sulfation in phylogenetically distant organisms with distinct bioluminescence systems. We use gene expression analyses, recombinant protein expression, and *in vitro* biochemical assays to identify and functionally test the five most highly expressed sulfotransferase genes from the light-producing organ of the cypridinid ostracod *Vargula tsujii* and report multiple sulfotransferases capable of sulfating vargulin *in vitro*. Our results, taken together with functional evidence of luciferin sulfotransferases previously published on sea pansies and fireflies [15,16] provide an example of an ancient and

pervasive gene family that was recruited multiple times independently for a similar purpose.

Materials and Methods

(a) Reference transcriptome assembly for Vargula tsujii

We constructed a reference transcriptome for *Vargula tsujii* by pooling RNA from 88 specimens across various instars (see table S1; figure S3, S4A). We extracted RNA using TRIzol(R) for sequencing at the BYU Sequencing Center using PacBio Sequel II ("IsoSeq") using two transcript size fractions: 4-10 kb selected using BluePippin (Sage Science) and a non-size-selected fraction. Next, we processed circular consensus sequencing reads with the IsoSeq v3 pipeline (https://github.com/PacificBiosciences/IsoSeq) and combined them with Illumina short reads (SRR21201581, [18]) using rnaSPAdes with flags --rna, --only-assembler, and --trusted-contigs [19]. We clustered sequences with ≥ 95% identity using cd-hit-est [20,21].

(b) Sulfotransferase candidates from the upper lip of Vargula tsujii

To quantify gene expression, we collected RNA from light organs (upper lips) of five male and four female ostracods maintained in culture [22] (figure S4B). After extracting RNA with Trizol, we used the Genomic Sequencing and Analysis Facility at UT-Austin for library preparations and Tag-Seq profiling using Illumina HiSeq 2500, SR100 [23] . We processed Tag-Seq reads (https://github.com/z0on/tag-based_RNAseq) by mapping to our reference transcriptome with Bowtie2 v.2.3.4.3 [24] (table S2). For protein sequence identification, we translated sequences with TransDecoder v.5.5.0 and used HMMER v.3.2.1 (hmmer.org) to find complete protein sequences containing a sulfotransferase domain (Pfam ID: PF00685) of e-value ≤ 1e-3.

(c) Cloning, expression, and purification of Vargula tsujii sulfotransferase candidates

We synthesized DNA sequences for candidate luciferin sulfotransferases ST1-5 (figure S5) as gBlocks™ (Integrated DNA Technologies) and cloned them into bacterial expression vectors (ST3 and ST4 into pQE80L; ST1, ST2, and ST5 in pET SUMO to improve solubility [25](figure S6)) using Gibson assembly (table S3). These vectors allow protein expression with N-terminal hexahistidine tags for purification using metal-affinity chromatography. We propagated plasmid constructs in *E. coli* NEB® 10-beta cells and verified by Sanger sequencing (Genewiz, South Plainfield, NJ).

We expressed proteins in *E. coli* BL21 cells via induction with IPTG, followed by incubating cultures at room temperature for 18 hours. We purified proteins by incubating lysed cell extracts with Ni-NTA agarose beads, then using gravity-flow metal-affinity

chromatography to elute hexahistidine-tagged proteins. We concentrated eluates by spin filtration (10 kDa molecular weight cutoff) and assessed protein purity via SDS-PAGE (figure S7, table S4) and densitometry. We quantified purified protein yields using a Bradford assay (Bio-Rad Laboratories, Hercules, CA) calibrated with a bovine serum albumin standard. These methods successfully purified all five proteins with > 90% homogeneity (table S5), and we prepared proteins freshly for each experiment.

(d) Sulfotransferase Functional Assays

We tested for sulfotransferase activity using two functional assays. First, we used a commercially available sulfotransferase assay (R&D Systems, Minneapolis, MN) that probes for the formation of 3'-phosphoadenosine-5'-phosphate (PAP), a byproduct of sulfation, by measuring the increase in absorbance at 620 nm following the enzymatic hydrolysis of PAP into phosphate, which binds to malachite green. In this assay, the phosphatase used to hydrolyze PAP into phosphate has some activity with PAPS, so we used an enzyme-free negative control and kept PAPS the same between experiments and control.

The second assay quantifies decreases in luminescence to test each candidate protein's ability to convert luciferin to its sulfated form, which luciferase does not oxidize. For ostracod luciferase, we transfected CHO cells with a pCMV *Cypridina* luciferase vector, and purified the protein with metal-affinity chromatography (figure S8). We first mixed each purified protein with ostracod luciferin and PAPS, incubated at 25 °C for 3 hours, then quenched reactions by heat-inactivation. After cooling, we added the diluted reaction product to purified ostracod luciferase and immediately measured bioluminescence. We compared luminescence between test and enzyme-free reactions with a decrease in light emission indicating less luciferin available for light production due to conversion to its sulfated form. In addition to luciferin sulfation assays, we used an established p-nitrophenol sulfation assay to evaluate sulfotransferase candidates for their ability to sulfate p-nitrophenol, a common substrate for many sulfotransferases, including those unrelated to bioluminescence (figure S9).

(e) Phylogenetic analyses

We estimated a phylogenetic tree for sulfotransferase candidates from *V. tsujii* and a tree of functionally tested sulfotransferases from the literature (as of March 2023) (table S6, Supplemental References). For aligning protein sequences, we used MAFFT (v. 7.429) [26], inferred maximum likelihood phylogenies using IQ-TREE (v. 1.6.12) [27], assuming the best-fit substitution model (LG+I+G4 for candidates, LG+R5 for functionally tested) according to Bayesian Information Criterion, as determined with ModelFinder [28], and performed ultrafast bootstrap approximation [29] with 1000 bootstrap replicates. To midpoint root the phylogeny and visualize bootstrap values and

expression values, we used the R packages phytools (v. 1.5-1) [30] and ggtree (v. 3.9.0) [31].

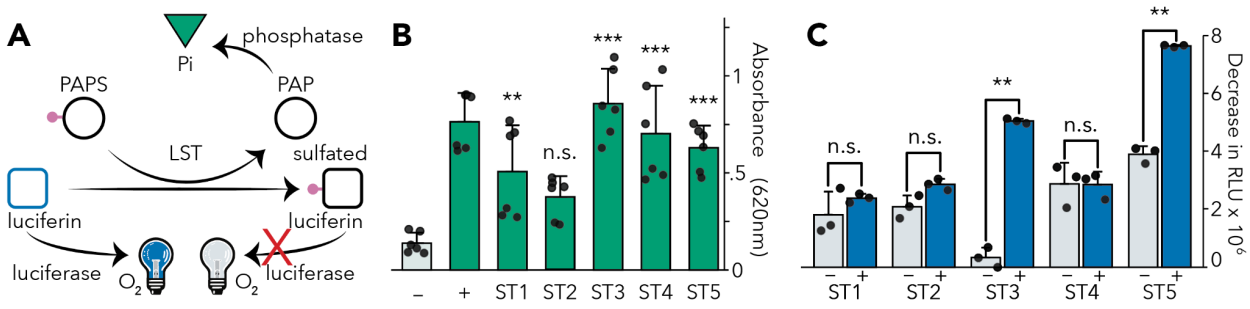

Figure 1. Based on two different assays, multiple candidate sulfotransferases from *Vargula tsujii*, especially ST3 and ST5, exhibit luciferin sulfation activity. **A**) Schematic of the chemistry underlying the two assays. **B**) Results of the absorbance-based sulfotransferase assay that probes for the conversion of PAPS (sulfate donor) to PAP. Our positive (+) control used 50 µL of 10 ng/µL Golgi-resident PAP-specific 3' phosphatase and 50 µM PAP to confirm phosphatase activity and a negative (–) control used 50 µL of 10 ng/µL coupling phosphatase, 100 µM luciferin, 200 µM lithium-free PAPS, but lacks any sulfotransferase. The test reactions used 50 µL of 10 ng/µL coupling phosphatase, 100 µM luciferin, and 200 µM lithium-free PAPS, and 10 µM sulfotransferase. Three genes showed the highest activity (ST3-5) and ST1 also showed significant activity under these conditions. **C**) Results of the bioluminescence assay that probes for a decrease in light emission due to sulfation of luciferin, which decreases luciferin available for oxidation. Negative control reactions (gray bars) lack PAPS. Test reactions (blue bars) included 10 µL of 10 µM ostracod luciferin and 100 uM PAPS with 1 µL of each purified protein (final concentrations ST1: 15 µM, ST2: 2 µM, ST3: 51 µM, ST4: 32 µM, ST5: 25 µM). Variations across the negative controls likely arise from the spontaneous oxidation of luciferin. Compared to control PAPS-free reactions, two candidates (ST3, ST5) exhibit a significant decrease in light production indicative of sulfation activity sufficient to modulate bioluminescence in vitro. ST2 may have some activity, but was not statistically significant with Welch's test.

Results

Functional assays of sulfotransferase genes from the light organ of Vargula tsujii

From Tag-seq on light-producing organs of *Vargula tsujii* (N = 9), we identified 40 genes, expressed in at least one individual, that contain domains with significant similarity to sulfotransferase domains, with nine expressed in all individuals. We selected the five (ST1-5) most highly expressed candidate sulfotransferases in light organs (figure S10) for recombinant expression and performed two functional assays to test for luciferin (vargulin) sulfation activity (figure 1b,c). Based on the commercial absorbance-based assay that probes for sulfotransferase-dependent conversion of PAPS to PAP, we found three candidates have the highest fold-change (5-6 fold) in activity (ST3, ST4, ST5) with ST1 also showing statistically significant change in absorbance under the conditions used (figure 1b). Based on the bioluminescence assay, which measures the decrease in light emission caused by sulfating luciferin, we infer the largest magnitude of luciferin sulfation in ST3 and ST5 (figure 1c).

Ostracods, fireflies, and sea pansies sulfate luciferin by using homologous genes

Genes of ostracods, fireflies, and sea pansies convergently evolved sulfation capabilities (figure 2), probably independently recruiting sulfotransferase genes as a

mechanism for metabolizing luciferin. Our phylogenetic analysis of functionally characterized luciferin sulfotransferases (LSTs) are most closely related to sulfotransferases of non-luminous animals. We find firefly LST to be most closely related to a gene from a non-luminous silk moth *Bombyx mori* that sulfates multiple substrates. The LST of the sea pansy *R. muelleri* is most closely related to sulfotransferase genes from a non-luminous tick (*Ixodes scapularis*) and a nematode (*Caenhorabditis elegans*). The five candidate LSTs from the ostracod *V. tsujii* we tested form a clade sister to genes from non-luminous *B. mori*, and *D. melanogaster*, and ST3 from firefly, which does not sulfate luciferin [15]. The close relationship between luciferin sulfotransferases to genes from non-luminous animals or genes lacking the ability to sulfate luciferin, strongly support convergent recruitment of LSTs in fireflies, sea pansies, and ostracods (figure 2).

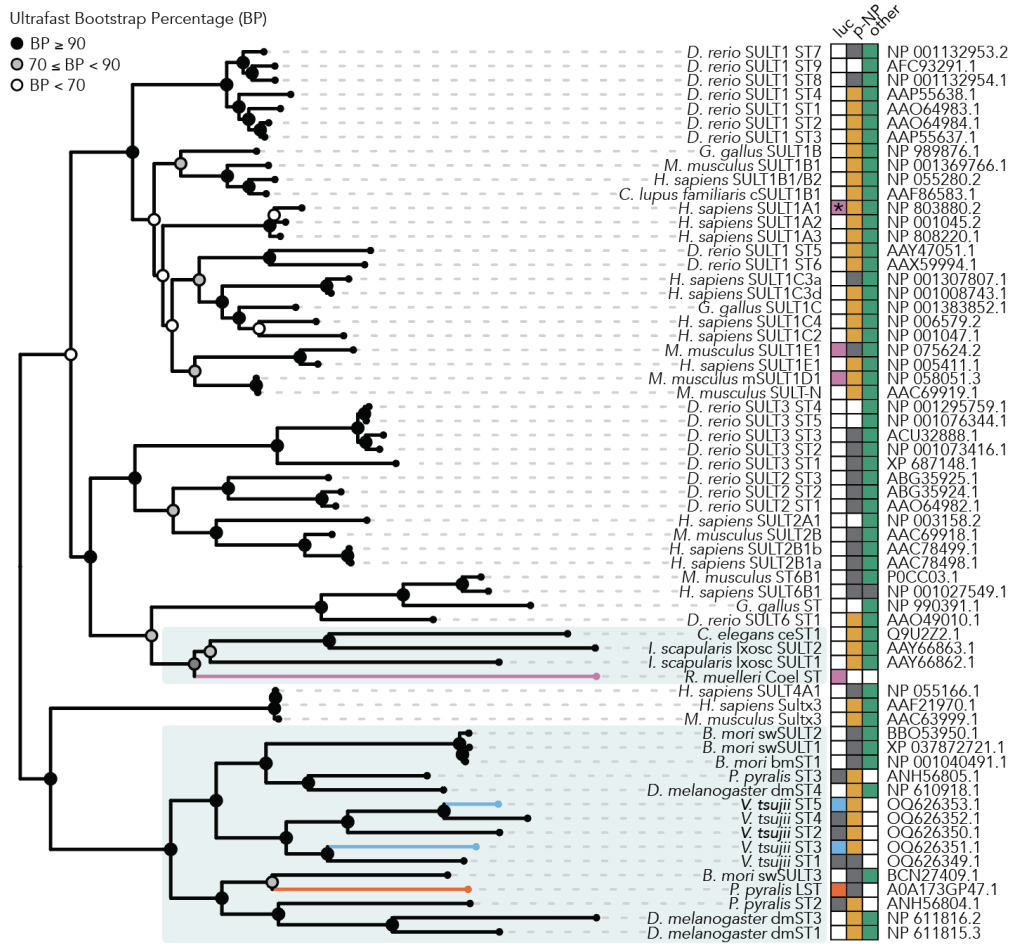

Figure 2. Maximum-likelihood phylogeny of functionally tested sulfotransferases. Luciferin sulfotransferases from luminous organisms are denoted by colored branches (firefly = red, ostracods = blue, sea pansies = purple). Invertebrate sulfotransferases are highlighted in gray. Colored squares (red = firefly luciferin, blue = ostracod luciferin for genes with strong activity in both assays herein, purple = sea pansy luciferin, orange = p-nitrophenol, green = other) indicate functionally demonstrated substrates for each sulfotransferase; gray squares indicate molecules that are not sulfated by the respective sulfotransferase; white squares are untested. *Sulfates the luciferin coelenterazine at a different site compared to sea pansy sulfated luciferin [32] Accession numbers are listed on the right. Ultrafast bootstrap values are represented by colored circles at each node.

Discussion

Convergently evolved traits may independently deploy members of deeply conserved gene families, a pattern called deep homology [1]. Although criticized for imprecision and incompleteness [33,34], the term deep homology highlights the importance of ancient evolutionary histories in shaping new traits [35]. Here we report sulfation in bioluminescence systems may have evolved by independently recruiting homologous genes from the ancient sulfotransferase gene family present across animals. We provide functional evidence for multiple sulfotransferases capable of sulfating luciferin in the ostracod *Vargula tsujii* and synthesize our results with those previously described in fireflies and sea pansies.

Although establishing the genetic basis of organismal functions requires gene knockdowns, multiple lines of evidence point to sulfotransferase involvement in the bioluminescence of ostracods. Our results demonstrate multiple sulfotransferase genes of *V. tsujii* sulfate luciferin, but determining how many using statistical significance depends on the assay (figure 1). The commercial assay that probes for PAP indicates four genes with statistically significant sulfation of luciferin (with ST3-5 showing especially strong differences between test and control) and the 'bioluminescence assay' that probes directly for luciferin sulfation indicates two genes (ST3 and ST5). The statistical difference between assays is likely due to specific experimental conditions, with the bioluminescence assay requiring a lower concentration of luciferin (10 μM) than the absorbance assay (100 μM), due to substrate inhibition of luciferase (figure S11). Nevertheless, results from both assays support multiple sulfotransferases highly expressed in the light organ — especially ST3 and ST5 — are capable of sulfating luciferin. Forthcoming work indicates ST3 is the only sulfotransferase with significant co-expression with luciferase across multiple stages and tissues (Oakley, biorxiv citable by the time of publication). Coupled with published results from crude protein extracts of *V. hilgendorfii* that sulfate luciferin [17], we hypothesize ST3 and possibly at least one other sulfotransferase (ST5) have important organismal functions in luminous ostracods, perhaps for use as a storage form of the easily oxidized luciferin.

A similar confluence of evidence suggests sulfotransferases are important to bioluminescence systems of fireflies and sea pansies. In fireflies, one sulfotransferase is expressed in lanterns and is able catalyze formation of sulfoluciferin from luciferin and PAPS, as confirmed recently [15]. This suggests an organismal function for a luciferin sulfotransferase in fireflies, again perhaps for creating a storage form of the substrate [15]. In the anthozoan *R. muelleri*, native sulfotransferase catalyzes sulfated luciferin into luciferin and the reverse reaction [36]. Recently, two sulfotransferase isoforms, similar in sequence to the native protein, were cloned, expressed, and shown to have sulfotransferase activity [16]. Because preparations from *Renilla* yield much higher amounts of luciferin sulfate compared to luciferin, the sulfated form probably acts as a storage form of the luminous substrate in the animal [36]. Therefore, despite difficulties

in knocking down specific sulfotransferases in ostracods, fireflies, or sea pansies, the combination of native preparations, gene expression studies, and heterologous expression suggest these animals use sulfotransferases in their respective bioluminescence systems.

Despite a growing realization of deep homology in evolutionary history, discovering homologous components in bioluminescence systems is arresting because of the distinctness of the other components. First, even the small-molecule substrates catalyzed by sulfotransferases are different: the luciferin of ostracods is vargulin, fireflies is D-luciferin, and sea pansy is coelenterazine. The luciferases also evolved from non-homologous gene families [10,11,37]. At the organismal level, bioluminescence is created by very different structures. Ostracod bioluminescence is secreted outside the body by the upper lip [38], firefly light is created in an abdominal structure called the lantern, and in sea pansies, light production occurs in specific cells called photocytes, located in the endoderm of two types of polyps. The presence of homologous sulfotransferases with similar functions along with non-homologous luciferases and distinct substrates provides a striking example of the often varied, cobbled-together elements of evolved complex systems. As more examples accumulate of cases of ancient genes independently recruited for similar functions, a logical pattern begins to emerge. Deeply homologous genes tend to have generally useful functions, such as sulfating small molecules in the case of sulfotransferases. Especially when there are few other solutions — as there is no other demonstrated mechanism for sulfation besides sulfotransferases – evolution may recruit the same gene family time and again, even in animals as distantly related as cnidarians and arthropods. This highlights how constraints — in this case determined by the number of biological solutions — may bias the options available to evolution [39,40].

Author Contributions

ESL: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data Curation, Writing - Original Draft and Review & Editing, Visualization. THO: Conceptualization, Resources, Writing - Review & Editing, Supervision, Project administration, Funding acquisition. AM: Methodology, Resources, Writing - Review & Editing, Supervision, Project administration, Funding acquisition. JAG: Software, Formal analysis, Investigation, Writing - Review & Editing, Resources, Data Curation. NTA: Methodology, Investigation. VLL: Investigation.

Data Availability

All raw sequencing data are available on NCBI (BioProject: PRJNA935772). Sulfotransferase sequences are available on GenBank (Accessions: OQ626349, OQ626350, OQ626351, OQ626352, OQ626353). Plasmid sequences, code for

analyzing and visualizing data, and data files for gene expression, phylogenetic analyses, and functional assays are available on Dryad (https://doi.org/10.25349/D94W5N).
For Anonymous Reviewers of this manuscript, the reviewer link is here: https://datadryad.org/stash/share/j_6734cHGfqb8Zt3RauJvtoY8ppssH08kliQkRwitZk

Funding information

AM: R35-GM133530 (NIH), DoD Peer-Reviewed Medical Research Program (PR191154). NTA: Chancellor's Fellowship (UCSB Graduate Division); THO & AM: Academic Senate Faculty Research Grant. ESL: NSF GRFP 1650114. JAG: BIO PRFB-1711201 (NSF). THO: ISO 1754770 (NSF). The authors acknowledge use of Biological Nanostructures Laboratory (led by J. Smith) within the California NanoSystems Institute, supported by the UCSB and UCOP. We used computational facilities purchased with funds from the National Science Foundation (CNS-1725797) and administered by the Center for Scientific Computing (CSC). The CSC is supported by the California NanoSystems Institute and the Materials Research Science and Engineering Center (MRSEC; NSF DMR 1720256) at UC Santa Barbara.

References Cited

1. Shubin N, Tabin C, Carroll S. 2009 Deep homology and the origins of evolutionary novelty. *Nature* **457**, 818–823.

2. Le Roy N, Jackson DJ, Marie B, Ramos-Silva P, Marin F. 2014 The evolution of metazoan α-carbonic anhydrases and their roles in calcium carbonate biomineralization. *Front. Zool.* **11**, 1–16.

3. Barott KL, Venn AA, Perez SO, Tambutté S, Tresguerres M. 2015 Coral host cells acidify symbiotic algal microenvironment to promote photosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 607–612.

4. Armstrong EJ, Roa JN, Stillman JH, Tresguerres M. 2018 Symbiont photosynthesis in giant clams is promoted by V-type H-ATPase from host cells. *J. Exp. Biol.* **221**. (doi:10.1242/jeb.177220)

5. Lau ES, Oakley TH. 2021 Multi-level convergence of complex traits and the evolution of bioluminescence. *Biol. Rev. Camb. Philos. Soc.* **96**, 673–691.

6. Peterson MK, Buck J. 1968 Light organ fine structure in certain Asiatic fireflies. *Biol.*

Bull. **135**, 335–348.

7. Spurlock BO, Cormier MJ. 1975 A fine structure study of the anthocodium in Renilla mülleri: Evidence for the existence of a bioluminescent organelle, the luminelle. *J. Cell Biol.* **64**, 15–28.

8. Huvard AL. 1993 Ultrastructure of the light organ and immunocytochemical localization of luciferase in luminescent marine ostracods (Crustacea: Ostracoda: Cypridinidae). *J. Morphol.* **218**, 181–193.

9. Tsarkova AS. 2021 Luciferins Under Construction: A Review of Known Biosynthetic Pathways. *Frontiers in Ecology and Evolution* **9**. (doi:10.3389/fevo.2021.667829)

10. Lorenz WW, McCann RO, Longiaru M, Cormier MJ. 1991 Isolation and expression of a cDNA encoding Renilla reniformis luciferase. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 4438–4442.

11. Oba Y, Ojika M, Inouye S. 2003 Firefly luciferase is a bifunctional enzyme: ATP-dependent monooxygenase and a long chain fatty acyl-CoA synthetase. *FEBS Lett.* **540**, 251–254.

12. Falany CN. 1997 Enzymology of human cytosolic sulfotransferases. *FASEB J.* **11**, 206–216.

13. Hori K, Nakano Y, Cormier MJ. 1972 Studies on the bioluminescence of Renilla reniformis. XI. Location of the sulfate group in luciferyl sulfate. *Biochim. Biophys. Acta* **256**, 638–644.

14. Nakamura M, Suzuki T, Ishizaka N, Sato J-I. 2014 Identification of 3-enol sulfate of Cypridina luciferin, Cypridina luciferyl sulfate, in the sea-firefly Cypridina (Vargula) hilgendorfii. *Tetrahedron* **70**, 2161–2168.

15. Fallon TR, Li F-S, Vicent MA, Weng J-K. 2016 Sulfoluciferin is Biosynthesized by a Specialized Luciferin Sulfotransferase in Fireflies. *Biochemistry* **55**, 3341–3344.

16. Tzertzinis G, Baker B, Benner J, Brown E, Corrêa IR Jr, Ettwiller L, McClung C, Schildkraut I. 2022 Coelenterazine sulfotransferase from Renilla muelleri. *PLoS One* **17**, e0276315.

17. Nakamura M, Suzuki T, Ishizaka N, Sato J-I, Inouye S. 2014 Identification of 3-enol sulfate of Cypridina luciferin, Cypridina luciferyl sulfate, in the sea-firefly Cypridina (Vargula) hilgendorfii. *Tetrahedron* **70**, 2161–2168.

18. Ellis EA *et al.* 2022 Sexual signals persist over deep time: ancient co-option of bioluminescence for courtship displays in cypridinid ostracods. *Syst. Biol.* (doi:10.1093/sysbio/syac057)

19. Bushmanova E, Antipov D, Lapidus A, Prjibelski AD. 2019 rnaSPAdes: a de novo

transcriptome assembler and its application to RNA-Seq data. *Gigascience* **8**. (doi:10.1093/gigascience/giz100)

20. Li W, Godzik A. 2006 Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659.

21. Fu L, Niu B, Zhu Z, Wu S, Li W. 2012 CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152.

22. Goodheart JA *et al.* 2020 Laboratory culture of the California Sea Firefly Vargula tsujii (Ostracoda: Cypridinidae): Developing a model system for the evolution of marine bioluminescence. *Sci. Rep.* **10**, 10443.

23. Meyer E, Aglyamova GV, Matz MV. 2011 Profiling gene expression responses of coral larvae (Acropora millepora) to elevated temperature and settlement inducers using a novel RNA-Seq procedure. *Mol. Ecol.* **20**, 3599–3616.

24. Langmead B, Salzberg SL. 2012 Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359.

25. Butt TR, Edavettal SC, Hall JP, Mattern MR. 2005 SUMO fusion technology for difficult-to-express proteins. *Protein Expr. Purif.* **43**, 1–9.

26. Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780.

27. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015 IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274.

28. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589.

29. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2017 UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518–522.

30. Revell LJ. 2012 phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223.

31. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017 Ggtree : An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36.

32. Inouye S, Matsuda K, Nakamura M. 2023 Enzymatic sulfation of coelenterazine by human cytosolic aryl sulfotransferase SULT1A1: identification of coelenterazine C2-benzyl monosulfate by LC/ESI-TOF-MS. *Biochem. Biophys. Res. Commun.* **665**, 133–140.

33. Scotland RW. 2010 Deep homology: a view from systematics. *Bioessays* **32**, 438–449.

34. DiFrisco J, Wagner GP, Love AC. 2023 Reframing research on evolutionary novelty and co-option: Character identity mechanisms versus deep homology. *Semin. Cell Dev. Biol.* **145**, 3–12.

35. Rose MR, Oakley TH. 2007 The new biology: beyond the Modern Synthesis. *Biol. Direct* **2**, 30.

36. Cormier MJ, Hori K, Karkhanis YD. 1970 Studies on the bioluminescence of Renilla reniformis. VII. Conversion of luciferin into luciferyl sulfate by luciferin sulfokinase. *Biochemistry* **9**, 1184–1189.

37. Oakley TH. 2005 Myodocopa (Crustacea: Ostracoda) as models for evolutionary studies of light and vision: multiple origins of bioluminescence and extreme sexual dimorphism. *Hydrobiologia* **538**, 179–192.

38. Abe K, Ono T, Yamada K, Yamamura N, Ikuta K. 2000 Multifunctions of the upper lip and a ventral reflecting organ in a bioluminescent ostracod Vargula hilgendorfii (Müller, 1890). *Hydrobiologia*

39. Losos JB. 2011 Convergence, adaptation, and constraint. *Evolution* **65**, 1827–1840.

40. Wake DB. 1991 Homoplasy: The Result of Natural Selection, or Evidence of Design Limitations? *Am. Nat.* **138**, 543–567.

Supplemental Material

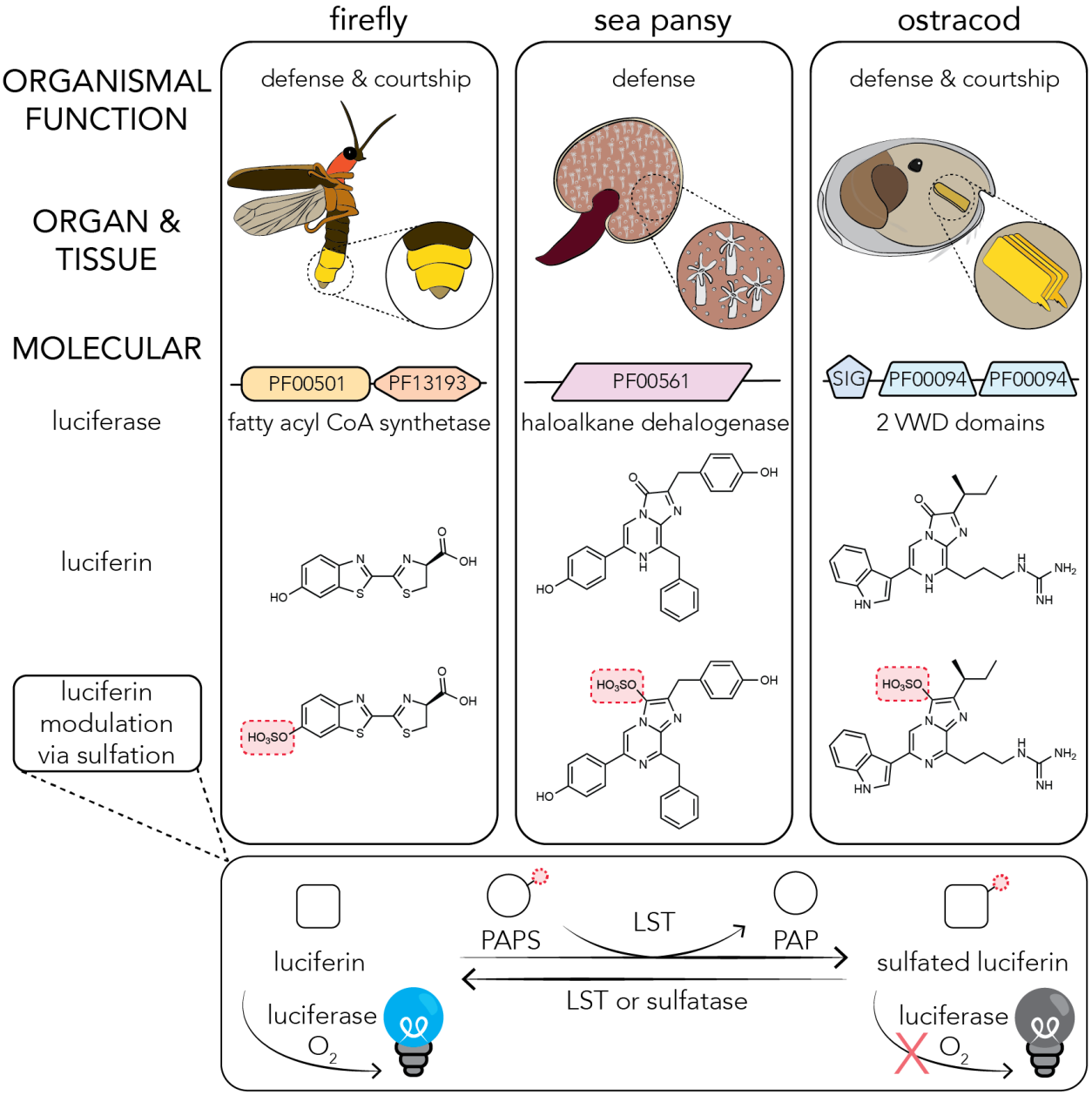

Figure S1. Convergently evolved bioluminescence systems of fireflies, sea pansies, and ostracods may vary in the extent of convergence across and within levels of biological organization.

These taxa use bioluminescence as a defense mechanism but produce light by using structurally different organs and tissues, non-homologous luciferases (PF00501 = AMP-dependent synthetase/ligase, PF13193 = AMP-binding enzyme C-terminal domain, PF00561 = alpha/beta hydrolase fold, SIG = signal peptide, PF00094 = von Willebrand factor

type D domain), and different luciferins. Despite these differences, they use sulfation as a common mechanism to modulate the chemical state of their luciferins. Luciferin sulfation is catalyzed by a luciferin sulfotransferase (LST), which transfers a sulfo group from the donor 3′-phosphoadenosine-5′-phosphosulfate (PAPS) to luciferin and produces 3’-phosphoadenosine-5’-phosphate (PAP) and sulfated luciferin, which cannot be oxidized by their respective luciferase to emit light. The sulfo group may be removed, either by a LST or sulfatase, to produce the active form of luciferin, which is oxidized by luciferase to produce light.

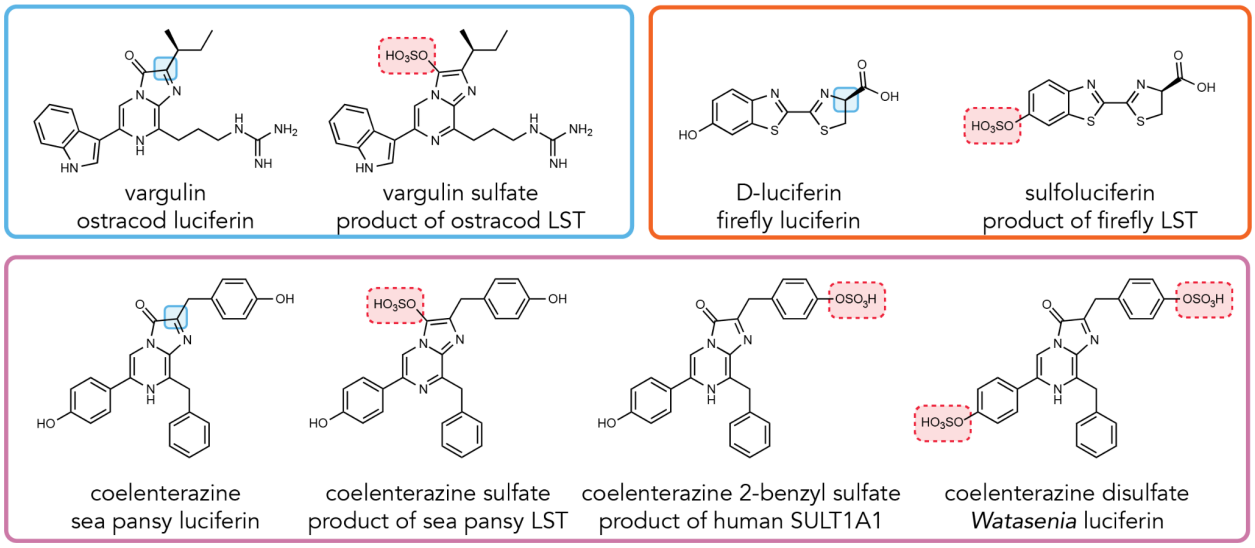

Figure S2. Chemical structures of luciferins from ostracods (vargulin), sea pansies (coelenterazine), and fireflies (D-luciferin).

Sites of luciferin modification by oxidation and sulfation are boxed in blue and red, respectively. The sites of sulfation in coelenterazine by sea pansy luciferin sulfotransferase (LST) and human SULT1A1 differ. The products of ostracod, firefly, and sea pansy LSTs are unable to react with their respective luciferases. However, in the firefly squid *Watasenia*, coelenterazine disulfate is used in its bioluminescence reaction.

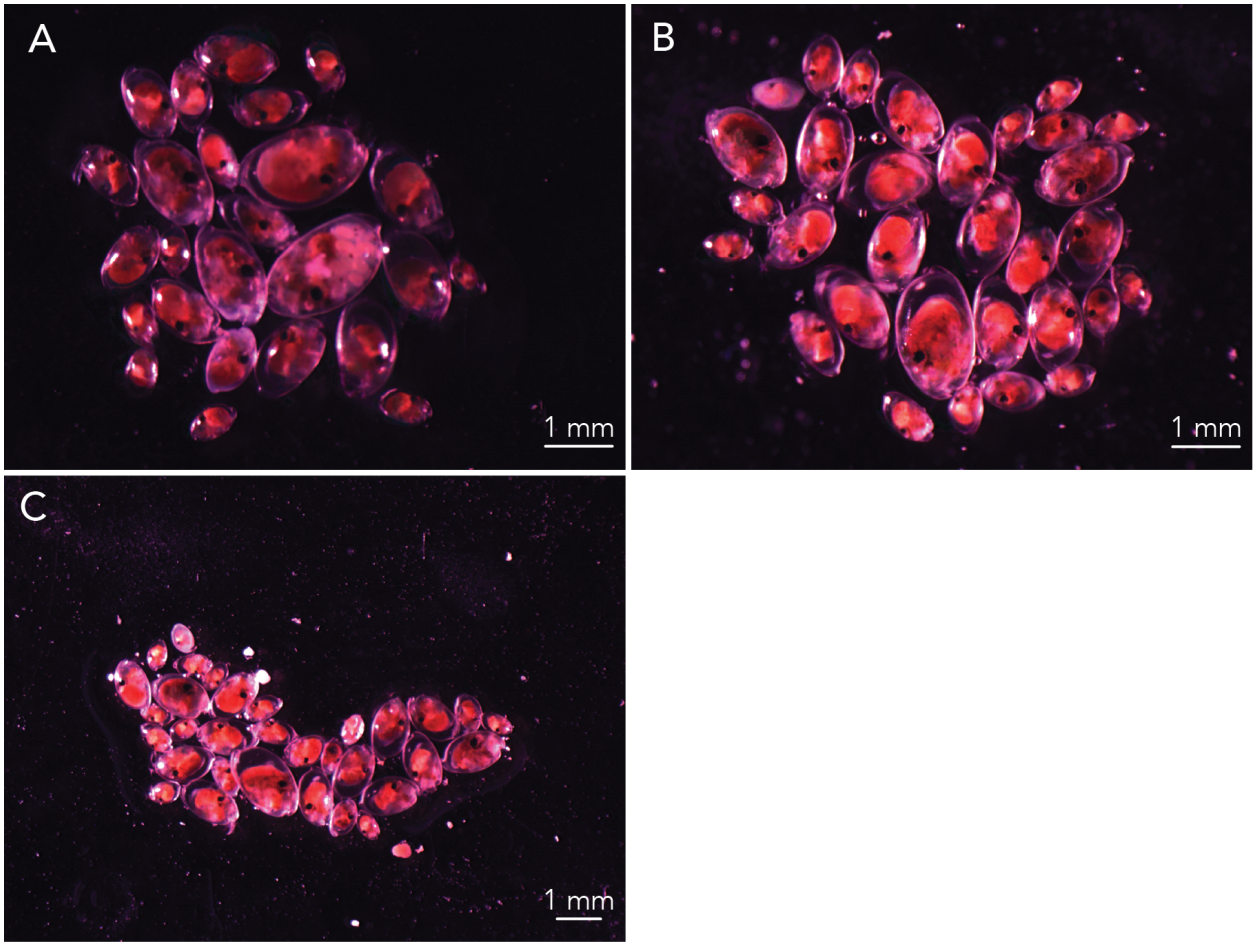

Figure S3. *Vargula tsujii* specimens sequenced via Iso-Seq to produce the reference transcriptome.

(A) Image “vt3”, 1.25x magnification. (B) Image “vt4”, 1.25x magnification. (C) Image “vt5”, 0.8x magnification. The names of each image are referred to in Supplemental Table 1, which lists the measurements and inferred instar stage of individuals in each photo.

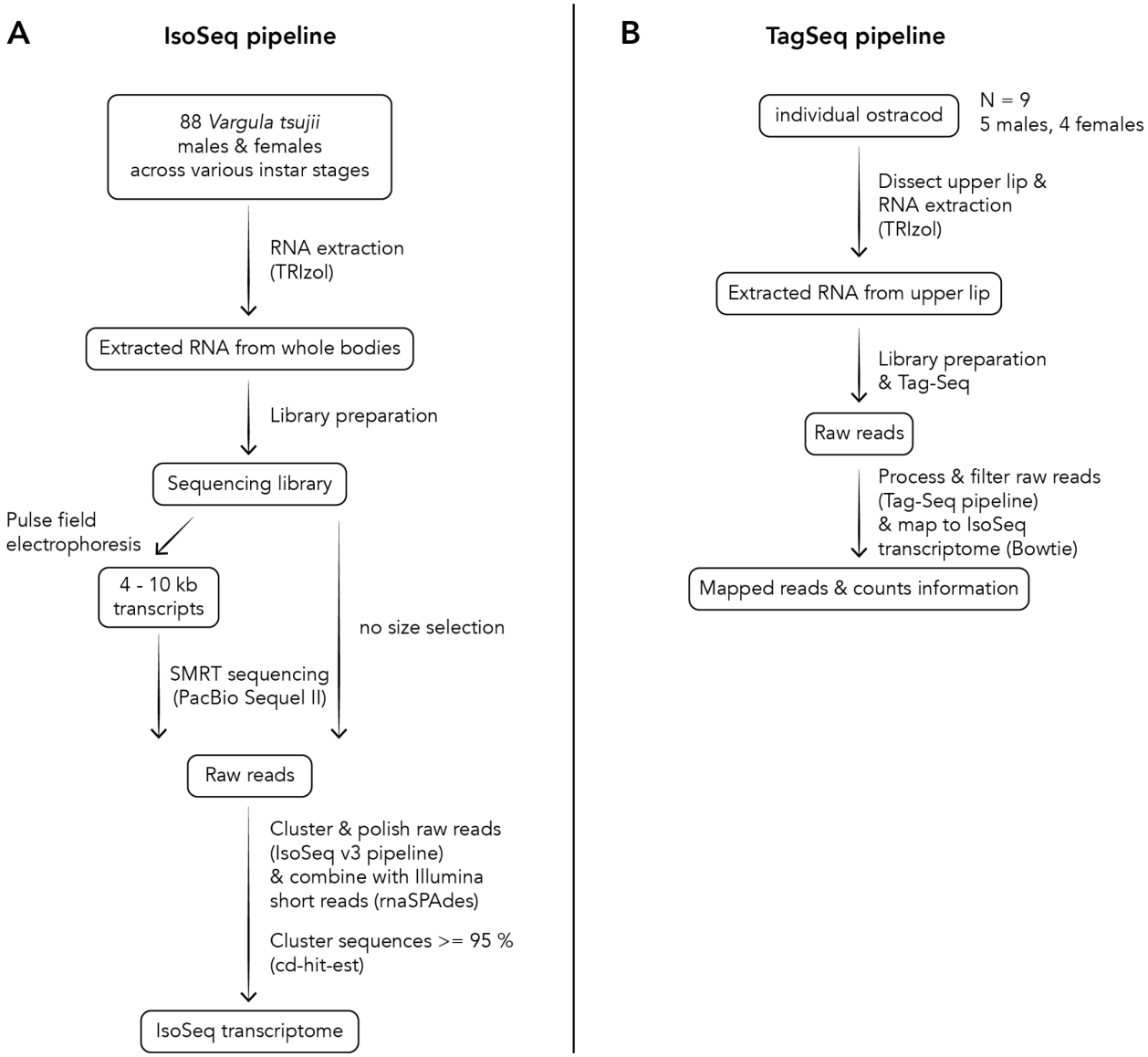

Figure S4. Flowchart of RNA sequencing and bioinformatics pipelines.

(A) We produced a high quality reference transcriptome by extracting and sequencing RNA from the whole bodies of 88 individuals using Single Molecular Real-Time (SMRT) Sequencing. (B) To characterize the expression of genes in the light-producing organ — the upper lip — of ostracods, we extracted and sequenced RNA from the upper lips of ostracods (N = 9) using Tag-Seq.

>ST1

MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGK
EMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGAAYPFTIEDYSSDLDPSPKVILQPGNWVMPNGFRPVAEK
IYNMEVREDDIWLLSWPKTGTTWISEMIWTLCHDSDTELAKTIPHFKRWAFPENDSMWADFSDYLRMQLSDGVPFPKEV
VHQLESDSFEDADKMEGPRFVRSHLPICMLPPNLLDTCKCIYMLRNPMDTCVSFYHFMKTVADVEIDIGMYINQFVADE
AYFSPLWAHVKEAWQKRSHPNMMFLYYEELQKDLPASIRKVSKFLNRPVTEEQVEKLTTHLHIDNMRQNPMINMDEMKA
TGRLRPDAMFLRKGVVGDWKNNLTPKMEMQLINWWNNNMRDCDIRFPGIQLKSRF*

>ST2

MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGK
EMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGKKVPFTYEDCKVKDDDGEYVRVKPGNFLLPATIRDYGEE
IYNMELREGDVILLSWPKTGTHWCSELIWNICNDMNIEEAKSKSMLDRWRFIDVPSDKKKHVCALELTDKPMGRIEKLL
DVEGPRYIISHLPLSLFPPDTLDKCKVVYVTRNPMDTIVSLYRFLELRPDENTDFPTFVNNFMHDEVVFSPYWEHVGEA
WKLKEHSNLFFLTFEEMKKDLRATIKNVAKFFNKEFSIEQIDTLVDHLSFENMQSNKSVNAEYLNKTGKIKQGGQFIRK
GAVGDFKNSFTPAMQSQMKNWIQTNNRGCDIKFQGVAM*

>ST3

MRGSHHHHHHMTTFPYKFEEYQGGVADAIKLNPGGWVMLSTYQEDAEKLYNFGARPDDIWLVGHPKTGTTWSSELLWTV
LHDADVEGARTIPQMDRICYPEYDFIIVKRDLADMVQECKKLGKEVPEKLLKESTKSHYQMAVEHPSPRLLRTGMCLEL
LSPKLLDTCKVVYVTRNPLDTCVSFYHFINSMEATKEHVDFQDFFQLFLADLVPFGPFWKHVKYAWEKRHHPNMLFLFY
EDMKADLPGVIKKVADFVGKSLTEEQIERVAQYLSFDEMKKNKMVNLDSEREMGVWKKDSDFVRKGIVGDYKNHMTPEM
IEQLRKWWKEHMEDCDIKFPGFDEL*

>ST4

MRGSHHHHHHMRIQTRCVISERNMNEIPFTLEEFTSEDDPRRKVRVQPINYVLFAECKELLTRIYNLKFRPGDIVLITT
PKSGTHWLIEMLWTMINNVDTEAAKKDVHWERAPMLEINHIDIDVVKELKEKRQNGEKLMPMEDFWVDYDSLKKFEALP
SPRVCFTHLPIPMLPPNMLEVCKVIYLTRNPADVICSYHQFMHDRCHLHVDFDTFFNMFLDGTVLFTPYWSSVKEGWKI
RNHPNALFLLYEEMAKDLRGCIEKISHHIQCPLSEEQIDKLQDHLSFENMKKNPAINMTPLAGTAFMKKGGNVVRKGKI
GDSKTTLSEEQNKRLKDWCEANREECDIKFPGSDF*

>ST5

MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGK
EMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGADLPFTLEEFEVAEESRLKVRVQPGNFTLYAGYKQLASR
IYNIKFRSGDVVLMSTPKSGTHWAAEMLWNIINQVDIEAAKKDLLFRRVPMLEISHIDAEKTEELKEKKAAGENLLPIE
ETVIFNDSIETFEKLPSPRICYTHLPFILLPPNILEVCKIVYITRNPADVIVSYHPFMAAGPNFDVNFENFFKMFLDGT
ALFTPYWSSVKEAWKRKDHPNMLFLLYEEMKQDLRGCIKNVSKHLQRPLTGEQVDRLEDHLSFESMKSNSSVNMKHLAG
SAFFSKDGQLCRKGQIGDSLNTLSDDQKKQMKEWCDANRGDCNITFPGLDQ*

```
>C.noctiluca_luciferase
MKTLILAVALVYCATVHCQDCPYEPDPPNTVPTSCEAKEGECIDSSCGTCTRDILSDGLCENKPGKTCCRMCQYVIECR
VEAAGWFRTFYGKRFQFQEPGTYVLGQGTKGGDWKVSITLENLDGTKGAVLTKTRLEVAGDIIDIAQATENPITVNGGA
DPIIANPYTIGEVTIAVVEMPGFNITVIEFFKLIVIDILGGRSVRIAPDTANKGMISGLCGDLKMMEDTDFTSDPEQLA
IQPKINQEFDGCPLYGNPDDVAYCKGLLEPYKDSCRNPINFYYYTISCAFARCMGGDERASHVLLDYRETCAAPETRGT
CVLSGHTFYDTFDKARYQFQGPCKEILMAADCFWNTWDVKVSHRNVDSYTEVEKVRIRKQSTVVELIVDGKQILVGGEA
VSVPYSSQNTSIYWQDGDILTTAILPEALVVKFNFKQLLVVHIRDPFDGKTCGICGNYNQDFSDDSFDAEGACDLTPNP
PGCTEEQKPEAERLCNSLFAGQSDLDQKCNVCHKPDRVERCMYEYCLRGQQGFCDHAWEFKKECYIKHGDTLEVPDECK
GPDYKDHDGDYKDHDIDYKDDDDKGPHHHHHH*
```

Figure S5. Gene sequences of recombinant proteins produced in this study.

Proteins were tagged with a hexa-histidine tag (shown in orange) to facilitate purification by immobilized metal affinity chromatography (IMAC) using Ni-NTA resin. To improve solubility, ST1, ST2, and ST5 were engineered to incorporate a SUMO tag (shown in blue) at the N terminus. Luciferase was further tagged with a 3 x FLAG epitope (shown in green) and hexa-histidine tag at the C terminus.

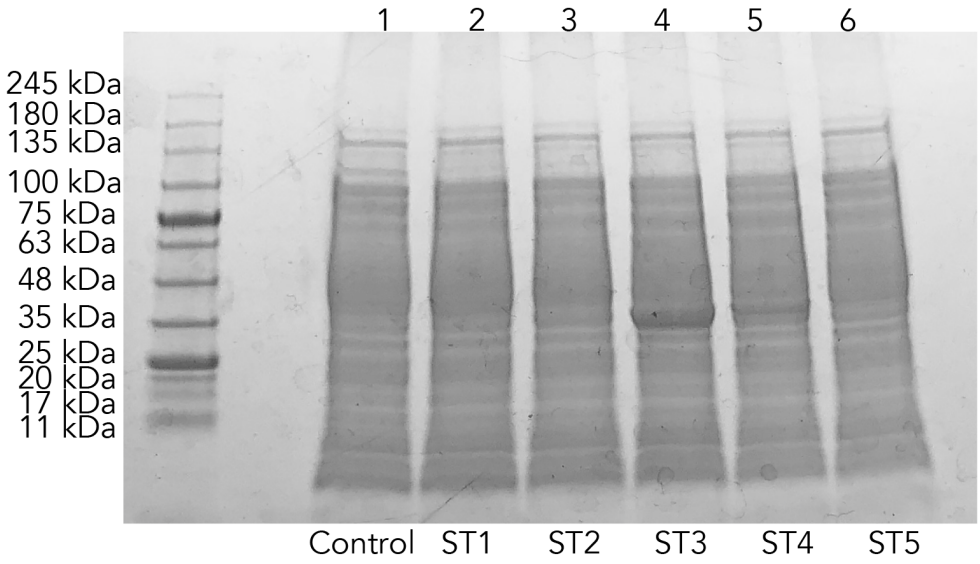

Figure S6. SDS-PAGE of crude lysates prepared from *E. coli* BL21 cells induced to express ST1-5.

Lane 1: cells transformed with control pQE80 vector lacking any ST, lane 2: ST1, lane 3: ST2, lane 4: ST3, lane 5: ST4, and lane 6: ST5 expressing *E. coli* cell lysates. SUMO-tags were not included for any ST construct. Bands at the expected size were detected only for ST3 (40.11 kDa) and ST4 (41.29 kDa).

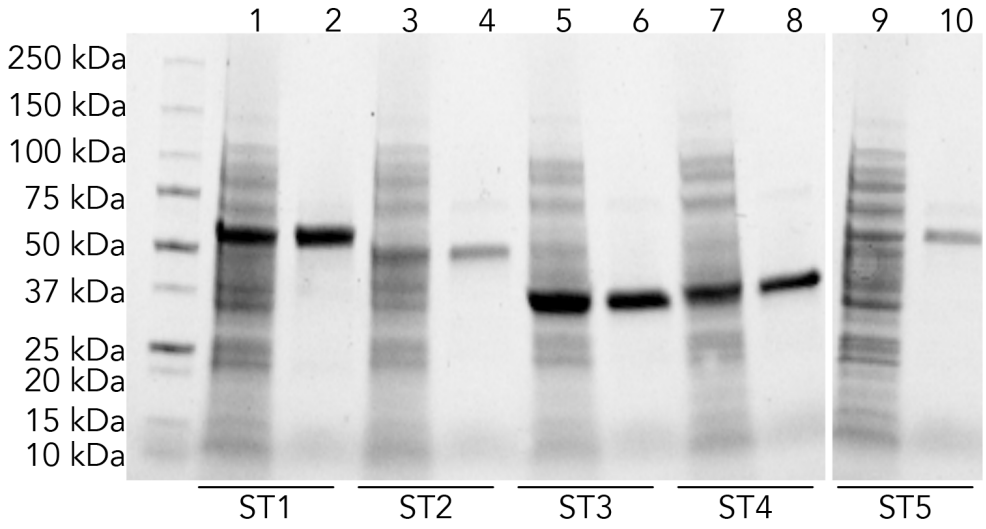

Figure S7. Polyacrylamide gel electrophoresis of purified sulfotransferase candidates.

Lanes 1, 3, 5, 7, and 9 contain (respectively) crude lysates prepared from ST1, ST2, ST3, ST4, and ST5 expressing *E. coli* cells. Lanes 2, 4, 6, 8, and 10 contain purified protein fractions of ST1, ST2, ST3, ST4, and ST5 prepared by IMAC with Ni-NTA resin. ST1, ST2, and ST5 are tagged with a SUMO peptide to increase solubility.

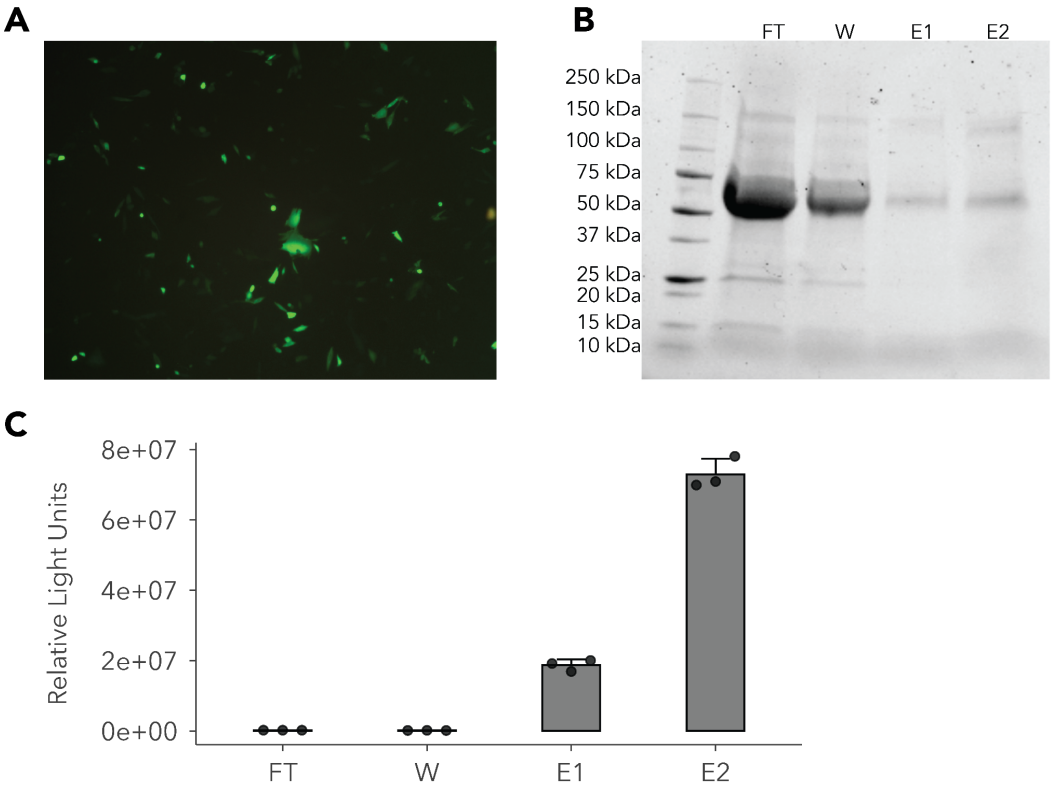

Figure S8. Expression and purification of ostracod luciferase.

(A) We co-transfected Chinese hamster ovary cells with the modified ostracod luciferase (from *Cypridina noctiluca*) vector and a GFP marker plasmid. We monitored expression of GFP in CHO cells to confirm transfection. Cells were excited with blue light and visualized using a 10X objective lens (Olympus CKX-41). 48 hours post transfection, we harvested the media, which contains the secreted luciferase, and purified the ostracod luciferase by performing IMAC using Ni-NTA agarose resin. (B) SDS-PAGE of each purification step. Lane 1: flow-through (FT); lane 2: washout fraction (W) from treatment with 10 mM imidazole containing buffer; lanes 3 and 4: elution fractions (E1 and E2) following treatment with 500 mM imidazole. (C) Presence of luciferase was verified by recording luminescence produced by incubating samples from each purification step with luciferin (N = 3).

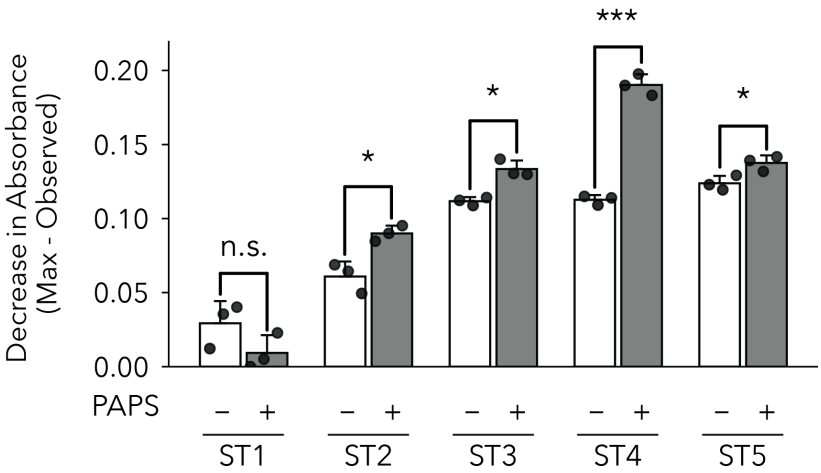

Figure S9. Four candidates, especially ST4, sulfate p-nitrophenol under the experimental conditions tested.

Sulfation of p-nitrophenol, a common substrate for many sulfotransferases, was assayed by measuring change in absorbance at 405 nm following a 15 hour incubation with purified sulfotransferase proteins in the absence and presence of PAPS. Control reactions, in white, contained 100 μL of 4 μM purified sulfotransferase and 100 μM p-nitrophenol but omitted PAPS to prevent sulfation activity. Experimental reactions, in gray, contained 100 μL of 4 μM purified sulfotransferase, 100 μM p-nitrophenol, and 500 μM PAPS. Four candidates — ST2, ST3, ST4, and ST5 — catalyze statistically significant p-nitrophenol sulfation under the conditions tested, with ST4 by far the highest effect size. Error bars represent the standard deviation (N = 3). * denotes $P < 0.05$, ** denotes $P < 0.01$, *** denotes $P < 0.001$, and n.s. denotes non-significance ($P \geq 0.05$). P values were calculated by performing Welch's t-tests between STs incubated with and without PAPS.

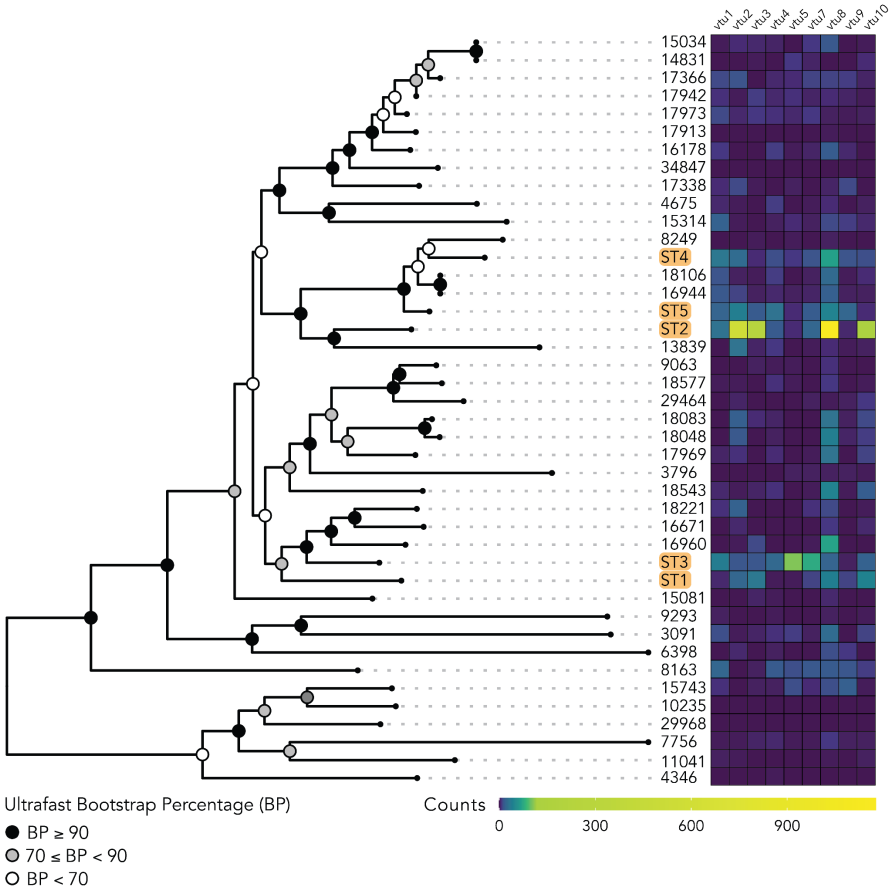

Figure S10. Phylogeny and gene expression of sulfotransferases from 9 individuals of *Vargula tsujii*.

A maximum-likelihood phylogeny was inferred using candidate sulfotransferases identified from the Iso-Seq reference transcriptome. Raw reads from Tag-seq were filtered and mapped to the reference transcriptome. The color bar represents read counts in each ostracod sample. Ultrafast bootstrap values are represented by colored circles at each node.

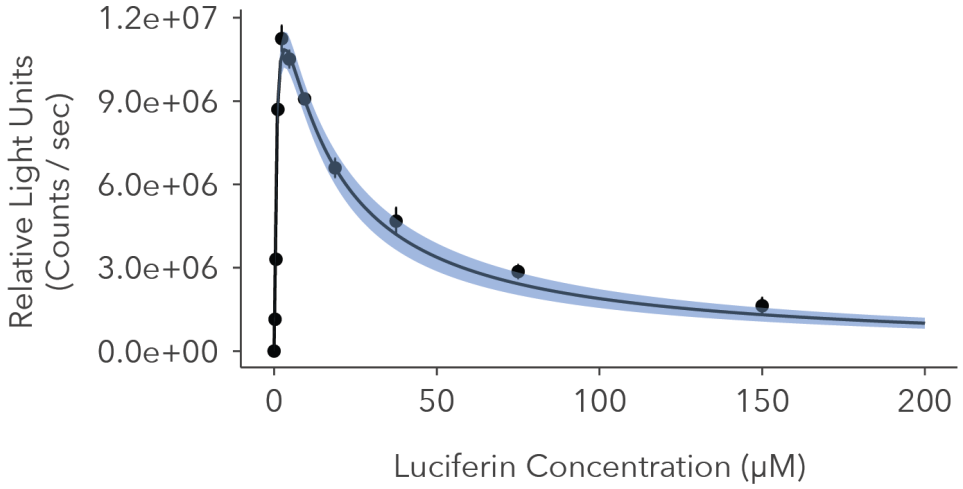

Figure S11. Ostracod luciferase exhibits substrate inhibition.

Bioluminescence produced by ostracod luciferase increases as substrate concentration increases — and peaks at a final substrate concentration of 1.5625 μM — but decreases at final substrate concentrations equal to or greater than 3.125 μM, suggesting ostracod luciferase exhibits substrate inhibition. We used a multi-channel pipette to manually add 65 μL of purified ostracod luciferase to 10 μL of luciferin, which resulting in a final concentration of 250 nM luciferase and varying final concentrations — 0 μM, 0.39 μM, 0.78125 μM, 1.5625 μM, 3.125 μM, 6.25 μM, 18.75 μM, 37.5 μM, 75 μM, and 150 μM — of luciferin. We incubated the reactions at room temperature for 12 minutes to stabilize the reading and quantified the luminescence using a microplate reader (Tecan Spark) with an integration time of 2 seconds. Data were fitted using a Michaelis-Menten model modified for substrate inhibition — $v = V_{max} [S] / K_m + [S] (1 + [S] / K_i)$ — where v = velocity (relative light units produced per second), V_{max} = maximum velocity, S = substrate concentration, K_m = Michaelis-Menten constant, and K_i = the dissociation constant for the inactive complex formed when two substrate molecules bind to an enzyme. Shaded areas represent 90 % confidence intervals. Estimated K_m = 0.8287 μM, V_{max} = 1.634e7, and K_i = 13.09 μM.

Supplemental Tables

Image	Animal	Length (mm)	Inferred stage
vt3	1	1.25	A-II
vt3	2	0.78	A-V
vt3	3	1.17	A-II
vt3	4	1.076	A-III
vt3	5	1.173	A-II
vt3	6	1.141	A-III
vt3	7	1.582	A-I/A Male
vt3	8	0.988	A-III
vt3	9	1.972	A Female
vt3	10	1.462	A-I
vt3	11	1.2	A-II
vt3	12	1.144	A-III
vt3	13	0.76	A-V
vt3	14	1.609	A-I/A Male
vt3	15	2.018	A Female
vt3	16	1.512	A-I
vt3	17	0.68	A-V
vt3	18	0.819	A-IV
vt3	19	1.244	A-II
vt3	20	0.769	A-V
vt3	21	1.064	A-III
vt3	22	1.403	A-I
vt3	23	1.504	A-I
vt3	24	0.738	A-V
vt3	25	0.792	A-IV
vt4	1	0.806	A-IV
vt4	2	0.954	A-IV
vt4	3	0.925	A-IV
vt4	4	1.502	A-I
vt4	5	1.398	A-I
vt4	6	1.35	A-I
vt4	7	1.405	A-I
vt4	8	1.352	A-I
vt4	9	0.701	A-V
vt4	10	0.761	A-V
vt4	11	0.997	A-III
vt4	12	0.815	A-IV
vt4	13	1.556	A-I/A Male
vt4	14	0.903	A-IV
vt4	15	0.727	A-V
vt4	16	1.257	A-II
vt4	17	1.434	A-I
vt4	18	1.457	A-I
vt4	19	1.073	A-III
vt4	20	1.445	A-I
vt4	21	1.082	A-III
vt4	22	1.482	A-I
vt4	23	1.985	A Female

vt4	24	1.416	A-I
vt4	25	1.365	A-I
vt4	26	0.915	A-IV
vt4	27	0.802	A-IV
vt4	28	0.943	A-IV
vt4	29	0.817	A-IV
vt4	30	0.957	A-IV
vt4	31	0.781	A-V
vt5	1	1.298	A-II
vt5	2	0.692	A-V
vt5	3	0.729	A-V
vt5	4	0.743	A-V
vt5	5	1.43	A-I
vt5	6	0.926	A-IV
vt5	7	0.77	A-V
vt5	8	0.85	A-IV
vt5	9	0.712	A-V
vt5	10	1.473	A-I
vt5	11	0.835	A-IV
vt5	12	1.415	A-I
vt5	13	1.449	A-I
vt5	14	0.94	A-IV
vt5	15	1.393	A-I
vt5	16	0.994	A-III
vt5	17	1.999	A Female
vt5	18	1.09	A-III
vt5	19	1.099	A-III
vt5	20	1.443	A-I
vt5	21	0.869	A-IV
vt5	22	0.723	A-V
vt5	23	1.414	A-I
vt5	24	0.952	A-IV
vt5	25	0.722	A-V
vt5	26	1.455	A-I
vt5	27	1.501	A-I
vt5	28	1.425	A-I
vt5	29	1.445	A-I
vt5	30	0.91	A-IV
vt5	31	1.581	A-I/A Male
vt5	32	0.817	A-IV

Table S1. Carapace measurements of *Vargula tsujii* specimens sequenced via Iso-Seq.

We sexed animals and for each specimen, measured the length of rostrum to keel. We inferred developmental stages based on the classification from [1].

Sample	Stage	Tissue	Raw Read Counts	Filtered Read Counts	% Raw Reads Kept	Mapped Reads	% Filtered Reads Mapped
vtu1	Adult Male	upper lip	3,784,061	1,210,393	32.0%	989,023	81.7%
vtu2	Adult Female	upper lip	2,508,701	607,251	24.2%	450,858	74.2%
vtu3	Adult Male	upper lip	3,286,003	1,166,599	35.5%	908,485	77.9%
vtu4	Adult Female	upper lip	3,785,756	1,094,550	28.9%	916,547	83.7%
vtu5	Adult Male	upper lip	2,845,340	1,021,284	35.9%	740,402	72.5%
vtu7	Adult Male	upper lip	5,725,267	2,098,711	36.7%	1,506,546	71.8%
vtu8	Adult Female	upper lip	2,725,196	653,138	24.0%	504,336	77.2%
vtu9	Adult Male	upper lip	3,486,104	1,211,298	34.7%	933,095	77.0%
vtu10	Adult Female	upper lip	4,071,147	1,197,806	29.4%	909,492	75.9%

Table S2. Tag-Seq sequencing, filtration, and mapping statistics.

Protein	Forward Gibson primer 5'-3'	Reverse Gibson primer 5'-3'
ST1	AGCTGAAGAGCCGTTTCTGAAGA CAAGCTTAGGTATTTATTCGGCG	ATGGTAAACGGATACGCAGCACC ACCAATCTGTTCTCTGTGAG
ST2	TCCAAGGGGTTGCTATGTAAAGA CAAGCTTAGGTATTTATTCGGCG	TAAGTAAATGGCACCTTTTTACC ACCAATCTGTTCTCTGTGAG
ST3	CAGGTTTCGATGAGCTTTAAAAG CTTAATTAGCTGAGCTTGGACTC	TTGTAGGGGAAAGTTGTCATGTG ATGGTGATGGTGATGCGAT
ST4	TCCCCGGCTCCGATTTCTGAAAG CTTAATTAGCTGAGCTTGGACTC	CAACGCGTCTGAATACGCATGTG ATGGTGATGGTGATGCGAT
ST5	TTCCCGGTTTGGACCAGTGAAGA CAAGCTTAGGTATTTATTCGGCG	AGCGTAAACGGAAGATCGGCACC ACCAATCTGTTCTCTGTGAG
Ostracod luciferase	GATTACAAGGATGACGATGACAA GGGCCCTCACCACCACCATCACC ACTGAGCGGCCGCAAAATCA	GATGTCATGATCTTTATAATCAC CGTCATGGTCTTTGTAGTCAGGG CCCTTGCACTCGTCGGGCA

Table S3. Oligonucleotide primers used in this work to clone the corresponding protein constructs (indicated in the left column) by Gibson assembly.

Protein	Molecular weight with purification/solubility tags (kDa)	Molecular weight without purification/solubility tags (kDa)
ST1	52.47	39.2
ST2	50.07	36.8
ST3	40.11	38.86
ST4	41.29	40.03
ST5	50.94	37.67
Ostracod luciferase	65.32 (with signal peptide) 63.39 (no signal peptide)	61.48 (with signal peptide) 59.54 (no signal peptide)

Table S4. Estimated molecular weights of recombinant proteins produced in this work.

Protein	Purity
ST1	98.76 ± 1.64 %
ST2	95.94 ± 1.93 %
ST3	98.61 ± 0.26 %
ST4	98.11 ± 0.15 %
ST5	93.01 ± 1.48 %

Table S5. Percentage purity of *V. tsujii* sulfotransferase candidates expressed in this work, as assessed by densitometry analysis of protein bands on a SDS-PAGE gel.

Each candidate was recombinantly expressed and purified twice. Densitometry analysis was performed for each purification and the percent purity was averaged. The values of percent purity reported in this table represent the average percent purity, and the + and -signs represent the maximum and minimum percent purity values, respectively, for each purification.

Organism	SULT name	Accession	Luciferin	p-nitrophenol	Other
Bombyx mori	bmST1	NP_001040491.1		No [2]	Yes [2]
Bombyx mori	swSULT1	XP_037872721.1		No [3]	Yes [3]
Bombyx mori	swSULT2	BBO53950.1		No [3]	Yes [3]
Bombyx mori	swSULT3	BCN27409.1		No [4]	Yes [4]
Caenorhabditis elegans	ceST1	Q9U2Z2.1		Yes [5]	Yes [5]
Canis lupus familiaris	cSULT1B1	AAF86583.1		Yes [6]	Yes [6]
Danio rerio	SULT1 ST1	AAO64983.1		Yes [7]	Yes [7]
Danio rerio	SULT1 ST2	AAO64984.1		Yes [7]	Yes [7]
Danio rerio	SULT1 ST3	AAP55637.1		Yes [7]	Yes [7]
Danio rerio	SULT1 ST4	AAP55638.1		Yes [7]	Yes [7]
Danio rerio	SULT1 ST5	AAY47051.1		Yes [8]	Yes [8]
Danio rerio	SULT1 ST6	AAX59994.1		Yes [9]	Yes [9]
Danio rerio	SULT1 ST7	NP_001132953.2		No [10]	Yes [10]
Danio rerio	SULT1 ST8	NP_001132954.1		No [10]	Yes [10]
Danio rerio	SULT1 ST9	AFC93291.1			Yes [11]
Danio rerio	SULT2 ST1	AAO64982.1		No [7]	Yes [7]
Danio rerio	SULT2 ST2	ABG35924.1		No [12]	Yes [12]
Danio rerio	SULT2 ST3	ABG35925.1		No [12]	Yes [12]
Danio rerio	SULT3 ST1	XP_687148.1		No [13]	Yes [13]
Danio rerio	SULT3 ST2	NP_001073416.1		No [13]	Yes [13]
Danio rerio	SULT3 ST3	ACU32888.1		No [14]	Yes [14]
Danio rerio	SULT3 ST4	NP_001295759.1			Yes [11]
Danio rerio	SULT3 ST5	NP_001076344.1			Yes [11]
Danio rerio	SULT6 ST1	AAO49010.1		Yes [15]	Yes [15]
Drosophila melanogaster	dmST1	NP_611815.3		Yes [16]	Yes [16]
Drosophila melanogaster	dmST3	NP_611816.2		Yes [16]	Yes [16]
Drosophila melanogaster	dmST4	NP_610918.1		Yes [16]	Yes [16]
Gallus gallus	SULT1C	NP_001383852.1		Yes [17]	Yes [17]
Gallus gallus	SULT1B	NP_989876.1		Yes [17]	Yes [17]

Gallus gallus	ST	NP_990391.1			Yes[18]
Homo sapiens	Sultx3	AAF21970.1		Yes [19]	Yes [19]
Homo sapiens	SULT1A1	NP_803880	Yes* [20]	Yes [21]	Yes [21]
Homo sapiens	SULT1A2	NP_001045		Yes [21]	Yes [21]
Homo sapiens	SULT1A3	NP_808220.1		Yes [21]	Yes [21]
Homo sapiens	SULT1B1/B2	NP_055280		Yes [22]	Yes [22]
Homo sapiens	SULT1C2	NP_001047.1		Yes [21]	Yes [21]
Homo sapiens	SULT1C3a	NP_001307807.1		No [23]	Yes [23]
Homo sapiens	SULT1C3d	NP_001008743.1		Yes [23]	Yes [23]
Homo sapiens	SULT1C4	NP_006579		Yes [21]	Yes [21]
Homo sapiens	SULT1E1	NP_005411		Yes [24]	Yes [24]
Homo sapiens	SULT2A1	NP_003158			Yes [21]
Homo sapiens	SULT2B1a	AAC78498.1		No [25]	Yes [25]
Homo sapiens	SULT2B1b	AAC78499.1		No [25]	Yes [25]
Homo sapiens	SULT4A1	NP_055166		No [26]	Yes [26]
Homo sapiens	SULT6B1	NP_001027549		No [26]	No [26]
Ixodes scapularis	Ixosc SULT 1	AAY66862.1		Yes [27]	Yes [27]
Ixodes scapularis	Ixosc SULT 2	AAY66863.1		Yes [27]	Yes [27]
Mus musculus	ST6B1	P0CC03.1		No [28]	Yes [28]
Mus musculus	Sultx3	AAC63999.1		Yes [19]	Yes [19]
Mus musculus	SULT2B	AAC69918.1		No [29]	Yes [29]
Mus musculus	SULT-N	AAC69919.1		Yes [29]	Yes [29]
Mus musculus	SULT1B1	NP_001369766.1		Yes [30]	Yes [30]
Mus musculus	SULT1E1	NP_075624.2	Yes [31]	No [32]	Yes [32]
Mus musculus	mSULT1D1	NP_058051.3	Yes [31]	Yes [33]	Yes [33]
Photinus pyralis	Ppyr LST	A0A173GP47.1	Yes [34]	No [34]	
Photinus pyralis	PpyrST2	ANH56804.1	No [34]	Yes [34]	
Photinus pyralis	PpyrST3	ANH56805.1	No [34]	Yes [34]	
Renilla muelleri	Coel ST		Yes [31]		
Vargula tsujii	VtST1	OQ626349	No	Undetected	

Vargula tsujii	VtST2	OQ626350	No	Small	
Vargula tsujii	VtST3	OQ626351	Yes**	Small	
Vargula tsujii	VtST4	OQ626352	No	Large	
Vargula tsujii	VtST5	OQ626353	Yes**	Small	

Table S6. Functionally characterized sulfotransferases and their ability to sulfate luciferins, p-nitrophenol, and other small molecule substrates.

Blank cells indicate missing or unknown data.

*Sulfates the luciferin coelenterazine at a different site compared to sea pansy sulfated luciferin, as shown in Figure S2.

**The ability to sulfate luciferin in VtST3 and VtST5 was supported by both the commercially available sulfotransferase assay and the luminescence assay.

Supplemental References

1. Goodheart JA *et al.* 2020 Laboratory culture of the California Sea Firefly Vargula tsujii (Ostracoda: Cypridinidae): Developing a model system for the evolution of marine bioluminescence. *Sci. Rep.* **10**, 1–15.

2. Kushida A, Horie R, Hattori K, Hamamoto H, Sekimizu K, Tamura H. 2012 Xanthurenic acid is an endogenous substrate for the silkworm cytosolic sulfotransferase, bmST1. *J. Insect Physiol.* **58**, 83–88.

3. Bairam AF, Kermasha ZW, Liu M-C, Kurogi K, Yamamoto K. 2020 Functional analysis of novel sulfotransferases in the silkworm Bombyx mori. *Arch. Insect Biochem. Physiol.* **104**, e21671.

4. Yamamoto K, Yamada N, Endo S, Kurogi K, Sakakibara Y, Suiko M. 2022 Novel silkworm (Bombyx mori) sulfotransferase swSULT ST3 is involved in metabolism of polyphenols from mulberry leaves. *PLoS One* **17**, e0270804.

5. Hattori K, Inoue M, Inoue T, Arai H, Tamura H-O. 2006 A novel sulfotransferase abundantly expressed in the dauer larvae of Caenorhabditis elegans. *J. Biochem.* **139**, 355–362.

6. Tsoi C, Falany CN, Morgenstern R, Swedmark S. 2001 Molecular cloning, expression, and characterization of a canine sulfotransferase that is a human ST1B2 ortholog. *Arch. Biochem. Biophys.* **390**, 87–92.

7. Suiko M, Sakakibara Y, Liu M-Y, Yang Y-S, Liu M-C. 2005 Cytosolic Sulfotransferases and Environmental Estrogenic Chemicals. *J. Pestic. Sci.* **30**, 345–353.

8. Yasuda S, Kumar AP, Liu M-Y, Sakakibara Y, Suiko M, Chen L, Liu M-C. 2005 Identification of a novel thyroid hormone-sulfating cytosolic sulfotransferase, SULT1 ST5, from zebrafish. *FEBS J.* **272**, 3828–3837.

9. Yasuda S, Liu C-C, Takahashi S, Suiko M, Chen L, Snow R, Liu M-C. 2005 Identification of a novel estrogen-sulfating cytosolic SULT from zebrafish: molecular cloning, expression, characterization, and ontogeny study. *Biochem. Biophys. Res. Commun.* **330**, 219–225.

10. Liu T-A *et al.* 2008 Identification and characterization of two novel cytosolic sulfotransferases, SULT1 ST7 and SULT1 ST8, from zebrafish. *Aquat. Toxicol.* **89**, 94–102.

11. Mohammed YI *et al.* 2012 Identification and characterization of zebrafish SULT1 ST9, SULT3 ST4, and SULT3 ST5. *Aquat. Toxicol.* **112-113**, 11–18.

12. Yasuda S, Liu M-Y, Yang Y-S, Snow R, Takahashi S, Liu M-C. 2006 Identification of novel hydroxysteroid-sulfating cytosolic SULTs, SULT2 ST2 and SULT2 ST3, from

zebrafish: Cloning, expression, characterization, and developmental expression. *Arch. Biochem. Biophys.* **455**, 1–9.

13. Yasuda T, Yasuda S, Williams FE, Liu M-Y, Sakakibara Y, Bhuiyan S, Snow R, Carter G, Liu M-C. 2008 Characterization and ontogenic study of novel steroid-sulfating SULT3 sulfotransferases from zebrafish. *Mol. Cell. Endocrinol.* **294**, 29–36.

14. Yasuda S *et al.* 2009 A novel hydroxysteroid-sulfating cytosolic sulfotransferase, SULT3 ST3, from zebrafish: identification, characterization, and ontogenic study. *Drug Metab. Lett.* **3**, 217–227.

15. Sugahara T, Liu CC, Govind Pai T, Liu MC. 2003 Molecular cloning, expression, and functional characterization of a novel zebrafish cytosolic sulfotransferase. *Biochem. Biophys. Res. Commun.* **300**, 725–730.

16. Hattori K, Motohashi N, Kobayashi I, Tohya T, Oikawa M, Tamura H-O. 2008 Cloning, expression, and characterization of cytosolic sulfotransferase isozymes from Drosophila melanogaster. *Biosci. Biotechnol. Biochem.* **72**, 540–547.

17. Wilson LA, Reyns GE, Darras VM, Coughtrie MWH. 2004 cDNA cloning, functional expression, and characterization of chicken sulfotransferases belonging to the SULT1B and SULT1C families. *Arch. Biochem. Biophys.* **428**, 64–72.

18. Cao H, Agarwal SK, Burnside J. 1999 Cloning and expression of a novel chicken sulfotransferase cDNA regulated by GH. *J. Endocrinol.* **160**, 491–500.

19. Sakakibara Y, Suiko M, Pai TG, Nakayama T, Takami Y, Katafuchi J, Liu MC. 2002 Highly conserved mouse and human brain sulfotransferases: molecular cloning, expression, and functional characterization. *Gene* **285**, 39–47.

20. Inouye S, Matsuda K, Nakamura M. 2023 Enzymatic sulfation of coelenterazine by human cytosolic aryl sulfotransferase SULT1A1: identification of coelenterazine C2-benzyl monosulfate by LC/ESI-TOF-MS. *Biochem. Biophys. Res. Commun.* **665**, 133–140.

21. Gamage N, Barnett A, Hempel N, Duggleby RG, Windmill KF, Martin JL, McManus ME. 2005 Human Sulfotransferases and Their Role in Chemical Metabolism. *Toxicol. Sci.* **90**, 5–22.

22. Wang J, Falany JL, Falany CN. 1998 Expression and characterization of a novel thyroid hormone-sulfating form of cytosolic sulfotransferase from human liver. *Mol. Pharmacol.* **53**, 274–282.

23. Kurogi K, Shimohira T, Kouriki-Nagatomo H, Zhang G, Miller ER, Sakakibara Y, Suiko M, Liu M-C. 2017 Human Cytosolic Sulphotransferase SULT1C3: genomic analysis and functional characterization of splice variant SULT1C3a and SULT1C3d. *J. Biochem.* **162**, 403–414.

24. Hempel N, Barnett AC, Bolton-Grob RM, Liyou NE, McManus ME. 2000 Site-directed mutagenesis of the substrate-binding cleft of human estrogen sulfotransferase. *Biochem. Biophys. Res. Commun.* **276**, 224–230.

25. Her C, Wood TC, Eichler EE, Mohrenweiser HW, Ramagli LS, Siciliano MJ, Weinshilboum RM. 1998 Human hydroxysteroid sulfotransferase SULT2B1: two enzymes encoded by a single chromosome 19 gene. *Genomics* **53**, 284–295.

26. Sun Y, Machalz D, Wolber G, Parr MK, Bureik M. 2020 Functional Expression of All Human Sulfotransferases in Fission Yeast, Assay Development, and Structural Models for Isoforms SULT4A1 and SULT6B1. *Biomolecules* **10**. (doi:10.3390/biom10111517)

27. Pichu S, Yalcin EB, Ribeiro JM, King RS, Mather TN. 2011 Molecular characterization of novel sulfotransferases from the tick, Ixodes scapularis. *BMC Biochem.* **12**, 32.

28. Takahashi S, Sakakibara Y, Mishiro E, Kouriki H, Nobe R, Kurogi K, Yasuda S, Liu M-C, Suiko M. 2009 Molecular Cloning, Expression and Characterization of A Novel Mouse SULT6 Cytosolic Sulfotransferase. *J. Biochem.* **146**, 399–405.

29. Sakakibara Y, Yanagisawa K, Takami Y, Nakayama T, Suiko M, Liu MC. 1998 Molecular cloning, expression, and functional characterization of novel mouse sulfotransferases. *Biochem. Biophys. Res. Commun.* **247**, 681–686.

30. Saeki Y, Sakakibara Y, Araki Y, Yanagisawa K, Suiko M, Nakajima H, Liu MC. 1998 Molecular cloning, expression, and characterization of a novel mouse liver SULT1B1 sulfotransferase. *J. Biochem.* **124**, 55–64.

31. Tzertzinis G, Baker B, Benner J, Brown E, Corrêa IR Jr, Ettwiller L, McClung C, Schildkraut I. 2022 Coelenterazine sulfotransferase from Renilla muelleri. *PLoS One* **17**, e0276315.

32. Song WC, Moore R, McLachlan JA, Negishi M. 1995 Molecular characterization of a testis-specific estrogen sulfotransferase and aberrant liver expression in obese and diabetogenic C57BL/KsJ-db/db mice. *Endocrinology* **136**, 2477–2484.

33. Shimada M, Terazawa R, Kamiyama Y, Honma W, Nagata K, Yamazoe Y. 2004 Unique properties of a renal sulfotransferase, St1d1, in dopamine metabolism. *J. Pharmacol. Exp. Ther.* **310**, 808–814.

34. Fallon TR, Li F-S, Vicent MA, Weng J-K. 2016 Sulfoluciferin is Biosynthesized by a Specialized Luciferin Sulfotransferase in Fireflies. *Biochemistry* **55**, 3341–3344.