Predictable genetic recruitment for luciferin sulfation in the convergent evolution of bioluminescence

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:** Emily Lau (emily.lau@lifesci.ucsb.edu); Arnab Mukherjee (arnabm@ucsb.edu); Todd Oakley (oakley@lifesci.ucsb.edu)

Abstract

Determinism and contingency shape evolutionary trajectories and the extent to which evolution is predictable depends on the relative contribution of each. However, the causes of predictable evolution are poorly understood. Here we take advantage of the fact that gene family recruitment may be more deterministic, and therefore predictable, when functional evolution is constrained to fewer genetic solutions. At the same time, the number of genes in a family capable of a specific function may be contingent on the vagaries of gene duplication mutations, functional changes, and maintenance, and therefore more difficult to predict. Here we analyze sulfotransferases using phylogenetics, gene expression, recombinant protein expression, and *in vitro* functional analyses. Sulfotransferases are the only enzymes known to catalyze sulfation, acting across a broad range of substrates and taxa, predicting that biological sulfation is catalyzed by sulfotransferases. We examine the bioluminescence system of ostracods and synthesize our findings with those reported previously for fireflies and sea pansies to show one aspect of bioluminescence, substrate metabolism, evolved convergently, each time recruiting sulfotransferase genes. However, our findings suggest that unlike fireflies, which have one known sulfotransferase gene specialized for substrate metabolism, ostracods contain several highly expressed sulfotransferase genes biochemically capable of metabolizing their bioluminescent substrate. While gene family recruitment may be predictable when limited genetic solutions exist, the number and specific identity of genes encoding a biological function may be difficult to predict because of the stochastic nature of gene duplication mutations and the diversity of mechanisms that fix and functionally differentiate these genes.

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

Introduction

Convergent evolution — the repeated evolution of similar phenotypes — provides replicates useful for investigating long-standing biological questions, such as to what extent evolution is predictable (Orgogozo 2015; Stayton 2015). Evolution may be more predictable and deterministic when similar features are shared across convergently evolved phenotypes, but less predictable and more contingent when similar features are not shared (Blount et al. 2018). Importantly, since many convergent phenotypes are complex, the extent of convergence may differ across levels of biological organization, or “multi-level convergent evolution” (Lau and Oakley 2020). Phenotypic convergence can be similar at one level of biological organization but radically different at other levels (Rosenblum et al. 2014). As one example, many vertebrates convergently evolved viviparity — the ability to produce live young — using structurally and functionally similar placentas, yet also using non-homologous genes (Foster et al. 2022). For viviparity, structural and functional evolution may be more predictable than genetic evolution. In some other cases, homologous genes are repeatedly co-opted to produce a similar phenotype — a special case of convergent evolution called “parallel evolution” (Rosenblum et al. 2014). For example, two squid species convergently evolved bioluminescence using structurally similar light-producing organs that express many genes homologous between the species. Therefore, squid light-producing organs largely evolved in parallel and show strong predictability of phenotypic evolution across

multiple levels of organization (Pankey et al. 2014). Fully understanding evolutionary predictability requires comparing convergent phenotypes across multiple biological levels of organization and identifying instances in which parallel genetic evolution underlies phenotypic convergence. Furthermore, understanding general rules of evolutionary predictability will require going beyond individual case studies to determine causes of these patterns — when and why is phenotypic evolution predictable at a genetic level?

For several reasons, bioluminescence — the biological production of light by a living organism via an intrinsic biochemical reaction or a symbiosis with luminous bacteria — provides an outstanding system for studying evolutionary predictability. First, bioluminescence is a complex trait that convergently evolved at least 94 times across the tree of life (Lau and Oakley 2020). Second, genes used to produce bioluminescence can be reliably tested via *in vitro* functional experiments (Shimomura 2019). Third, across origins of bioluminescence the extent of evolutionary convergence at different levels of biological organization is variable, leading to great diversity in form and function (Haddock et al. 2010; Widder 2010). Bioluminescence is used for many organismal functions: deterring predation, attracting prey, camouflaging, intraspecific communication, or courtship (Haddock et al. 2010). The production of bioluminescence results from a combination of several physiological functions — luminous organisms must biosynthesize (Kato et al. 2004; Kotlobay et al. 2018; Bessho-Uehara, Huang, et al. 2020) or externally obtain bioluminescent molecules (Thompson et al. 1988; Haddock et al. 2001; Bessho-Uehara, Yamamoto, et al. 2020), produce and control light

production (Krönström and Mallefet 2010; Duchatelet et al. 2020), and in some cases, modify the emitted light using filters (Denton et al. 1985). These physiological functions may be performed intracellularly or extracellularly using a variety of complex structures (Harvey 1952), including glandular organs that may secrete (Dilly and Herring 1978; Takenaka et al. 2012) or house bioluminescence (Morin 1983), modified fin rays that house luminous bacteria (Hulet and Musil 1968), or luminous tubules and ducts (Thacker and Roje 2009) that transport or house bioluminescent molecules. At a molecular level, light is produced when enzymes called luciferases oxidize small molecule substrates called luciferins (Shimomura 2019). To harness bioluminescence, luminous organisms may modulate luciferase-luciferin reactions by housing luciferases and luciferins in different cellular or subcellular locations (Clarke et al. 1962; Huvard 1993), or inactivating luciferins by using proteins to bind to or biochemically modify them (Lee et al. 1993; Inouye 2007; Nakamura et al. 2014; Fallon et al. 2016; Falaschi et al. 2019; Viviani et al. 2020). However, compared to knowledge of the function of light production and luciferase enzymes, much less is known about how luciferins are produced and metabolized for storage, transport, or modulation of light. Understanding how these physiological functions repeatedly evolved across convergently evolved bioluminescence systems may help elucidate evolutionary predictability.

Fireflies, sea pansies, and ostracods convergently evolved bioluminescence as a defense mechanism, which was later co-opted for courtship in fireflies (Powell et al. 2022) and some ostracods (Ellis et al. 2022). These three groups of organisms produce light autogenically by using structurally diverse organs or tissue (Peterson and Buck

1968; Spurlock and Cormier 1975; Huvard 1993), different luciferins (Tsarkova 2021), and non-homologous luciferases (Lorenz et al. 1991; Oba et al. 2003) (Figure 1). Despite these differences at a molecular level, these taxa convergently evolved the same biochemical mechanism, sulfation — or the transfer of sulfate (SO_3^-) from the donor 3′-phosphoadenosine-5′-phosphosulfate (PAPS) to a substrate (Falany 1997) — for modulating the chemical state of their luciferins (Hori et al. 1972; Nakamura et al. 2014; Fallon et al. 2016; Tzertzinis et al. 2022) (Figure 1, Supplemental Figure S1). Once sulfated, these luciferins are unable to react with their respective luciferases (Fallon et al. 2016), and in the cases of sea pansies and ostracods, sulfation prevents non-specific oxidation of their luciferins (Cormier and Kazuo 1964; Nakamura et al. 2014). Sulfation is catalyzed by a family of enzymes called sulfotransferases that contain a sulfotransferase domain with two highly conserved motifs necessary for binding to PAPS and transferring sulfate to a substrate (Driscoll et al. 1995). Sulfotransferases may exhibit “substrate promiscuity”, which we define as an enzyme’s ability to catalyze the same biochemical reaction in more than one substrate. Sulfotransferase genes underlying luciferin metabolism have been identified and functionally tested in fireflies (Fallon et al. 2016) and recently in sea pansies (Tzertzinis et al. 2022), but not in ostracods. Previous work did identify the presence of sulfated luciferin in the ostracod *Vargula hilgendorfii* (Nakamura et al. 2014). Nakamura et al. also found crude protein extracts from the whole organism could use PAPS to enzymatically sulfate luciferin to produce an inactive form of luciferin that could not react with luciferase and was less vulnerable to non-specific oxidation, suggesting the presence of a luciferin sulfotransferase. Additionally, crude protein extracts of *V.*

hilgendorfii can remove sulfate from sulfated luciferin to produce active luciferin (Nakamura et al. 2014). Since crude protein extracts from *V. hilgendorfii* can reversibly sulfate luciferin, this suggests sulfation may serve as a mechanism for luciferin metabolism in the organism (Nakamura et al. 2014). However, the genetic sequences of functional luciferin sulfotransferases in ostracods have yet to be identified.

Here, we identify sulfotransferase genes capable of sulfating luciferin *in vitro* from the cypridinid ostracod *Vargula tsujii*. Our results, taken together with functional evidence of luciferin sulfotransferases previously published on sea pansies and fireflies (Fallon et al. 2016; Tzertzinis et al. 2022) provide empirical evidence supporting independent and predictable recruitment of paralogous genes from a single gene family — sulfotransferases — and the parallel evolution of luciferin metabolism in three phylogenetically distant organisms with distinct luciferin substrates. We hypothesized that, similar to fireflies, the light-producing organ of ostracods contains a highly expressed sulfotransferase capable of sulfating ostracod luciferin. To test our hypothesis, 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 found in the light-producing organ of the cypridinid ostracod *Vargula tsujii*. We perform two types of functional tests to show that, surprisingly, multiple candidates are capable of catalyzing luciferin sulfation activity as evidenced by at least one assay, suggesting the number of luciferin sulfotransferase genes may vary across convergently evolved bioluminescence systems. Broadly speaking, while gene family recruitment for a particular function may be more deterministic and predictable

when genetic solutions are limited, the specific genes and number of genes encoding biochemical functions may be contingent on gene duplication, maintenance, and functional changes, and therefore more difficult to predict.

Results

Many sulfotransferase genes are expressed in the light-producing organ of Vargula tsujii

We used Tag-seq to profile gene expression in the upper lip (the light-producing organ) of *Vargula tsujii* (Supplemental Figure S3) and identified 40 sulfotransferase genes that were expressed in at least one *V. tsujii* individual. Of these 40 sulfotransferase genes, nine were expressed across all *V. tsujii* individuals sequenced (N = 9). To identify and functionally verify luciferin sulfotransferase genes in the ostracod *V. tsujii*, we hypothesized that, similar to fireflies (Fallon et al. 2016), the most highly expressed sulfotransferase in the light-producing organ will be capable of sulfating luciferin in a PAPS-dependent manner. We selected the five most highly expressed sulfotransferase genes in the light organ (Figure 2, Supplemental Figure 4), hereafter referred to as ST1-5, to recombinantly express and test for luciferin sulfation function.

Multiple sulfotransferases exhibit significant luciferin sulfation activity in vitro as evidenced by at least one assay

We show the luminous ostracod *V. tsujii* expresses multiple sulfotransferases capable of sulfating luciferin *in vitro*. To test whether there are luciferin sulfotransferases in *V. tsujii*, we performed two types of functional tests — a “universal sulfotransferase assay” and “luciferase assay” — for assessing luciferin sulfation activity, tested for significant

activity as compared with control reactions, and then corrected for multiple comparisons.

The universal sulfotransferase assay showed four candidate sulfotransferase genes — ST1, ST3, ST4, ST5 — exhibit significant sulfation activity on luciferin compared to the negative control, which lacked sulfotransferase (ST1: $p = 3.92e-3$, ST2: $p = 9.363e-2$, ST3: $p = 3.3e-7$, ST4: $p = 1.3e-5$, ST5: $p = 1.2e-4$, $N = 6$, Dunnett's test) (Figure 3A). This assay quantifies formation of phosphate from 3'-phosphoadenosine-5'-phosphate (PAP) — a metabolic by-product of sulfotransferases — as a proxy for sulfation activity. By avoiding a buildup of PAP in the reaction volume, this assay also mitigates potential product inhibition, which could limit sulfation activity *in vitro*.

The luciferase assay identified three of five candidate luciferin sulfotransferases — ST2, ST3, and ST5 — to have significant decreases in bioluminescence compared to PAPS-free controls (ST1: $p = 2.91e-1$, ST2: $p = 3.79e-2$, ST3: $p = 2.05e-5$, ST4: $p = 9.6e-1$, ST5: $p = 2.16e-5$, $N = 3$, unpaired 2-sample Student's t-test) (Figure 3B), suggesting the activities of ST2, ST3, and ST5 are sufficient to modulate bioluminescence *in vitro* because once sulfated, luciferins are unable to react with their corresponding luciferases to produce light. To account for testing multiple candidate sulfotransferases — which increases the probability of finding a functional luciferin sulfotransferase as a result of a type one error — we performed the Benjamini-Hochberg procedure and found only ST3 and ST5 to be statistically significant.

The discrepancy between the results of the universal sulfotransferase assay and luciferase assay could be explained by experimental differences between the two assays. The universal sulfotransferase assay required a high concentration of luciferin substrate (100 μM) to generate enough product to produce detectable absorbance measurements. In the luciferase assay, the concentration of luciferin (10 μM) was ten times less than that of the sulfotransferase assay. We found ostracod luciferase exhibits substrate inhibition (Supplemental Figure S7), which limited the concentration of luciferin that could be used in this assay. Since the rate of sulfation reactions depends on substrate concentrations (Atkinson 1966; James 2021), the differences in substrate concentrations between the two assays may have led to the observed statistical differences in product formation of sulfated luciferin. Inhibition of sulfotransferases by PAP (Chapman et al. 2004), a by-product of sulfation produced in both assays but removed only in the universal sulfotransferase assay, may have also contributed to the statistical differences in sulfation. Nevertheless, results from both assays support the presence of multiple sulfotransferases capable of sulfating luciferin and modulating bioluminescence in the light-producing organ of *V. tsujii*. Future research on the *in vivo* conditions of bioluminescence — such as concentrations of luciferin, luciferase, PAPS, and PAP — may further refine these *in vitro* assays by confirming physiologically relevant testing conditions. Additionally, future tests on the kinetic activities and biochemical characteristics of these sulfotransferases may provide further insight into how many sulfotransferase genes are used *in vivo*.

Four candidates exhibit significant sulfation activity with p-nitrophenol

Four candidates — ST2, ST3, ST4, and ST5 — catalyzed sulfation of p-nitrophenol relative to PAPS-free controls (ST1: $p = 1.45e-1$, ST2: $p = 1.17e-2$, ST3: $p = 4.22e-3$, ST4: $p = 6.91e-5$, ST5: $p = 2.92e-2$, $N = 3$, unpaired 2-sample Student's t-test) (Figure 4). Since many sulfotransferases sulfate a range of small molecules, we hypothesized ST1-5 may sulfate small molecules besides luciferin. To test this for ST1-5, we examined their ability to catalyze sulfation of p-nitrophenol, a xenobiotic substrate commonly sulfated by cytosolic sulfotransferases. These results support substrate promiscuity in the sulfotransferase activities of ST2, ST3, ST4, and ST5.

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

Ostracods, fireflies, and sea pansies convergently evolved sulfation as a mechanism for metabolizing luciferin by independently recruiting sulfotransferase genes (Figure 5). Phylogenetic analysis of functionally characterized sulfotransferases place the luciferin sulfotransferases of ostracods and fireflies in a clade of arthropod and mammalian sulfotransferases. This clade contains a mixture of sulfotransferases that may exhibit substrate promiscuity or specificity towards endogenous and xenobiotic small molecules, such as p-nitrophenol. Similarly, the luciferin sulfotransferase of sea pansies falls in a separate clade, which also contains sulfotransferases with substrate promiscuity or specialized activity towards substrates such as hormones and xenobiotic substrates.

Discussion

Convergently evolved phenotypes often result from a mixture of similar and different features across and within levels of biological organization (Woodard et al. 2011; Van Belleghem et al. 2023), providing excellent opportunities to explore general rules of evolutionary predictability. Here, we examine the convergently evolved bioluminescence systems of ostracods, fireflies, and sea pansies. We provide functional evidence for multiple sulfotransferases capable of sulfating luciferin in the ostracod *Vargula tsujii*, and synthesize these results with those previously described in fireflies and sea pansies, to illustrate how luciferin sulfation convergently evolved by recruiting paralogous genes from the same gene family. However, while luciferin sulfation evolved in parallel, light production convergently evolved by recruiting non-homologous genes (Lorenz et al. 1991; Oba et al. 2003; Oakley 2005). These differences between sulfation recruiting homologous genes and light production recruiting completely different genes raises the question: How might biochemical function inform the evolutionary predictability of complex biological systems like bioluminescence?

As a general rule, molecular evolution may be more predictable when biochemical functions have fewer possible genetic solutions (Christin et al. 2010; Ujvari et al. 2015), and less predictable when genetic redundancy permits many solutions (Berens et al. 2015; Foster et al. 2022). To test whether this rule may be applicable to the evolution of luciferin sulfation in ostracods, we hypothesized that, since sulfation in biological systems is genetically constrained to the sulfotransferase gene family (Coughtrie 2002), ostracods genetically encode luciferin sulfotransferases in their light producing organ. Our finding of multiple luciferin sulfotransferase genes, in combination with work

previously published on fireflies (Fallon et al. 2016) and sea pansies (Tzertzinis et al. 2022), support this general rule of evolutionary predictability and suggests that the evolution of luciferin sulfation may be predictable due to limited genetic solutions. In contrast, these three taxa use non-homologous luciferases to produce bioluminescence (Lorenz et al. 1991; Oba et al. 2003; Oakley 2005). The diversity of gene families that have been recruited during luciferase evolution suggest that enzyme-catalyzed light production can be accomplished by many genetic solutions and as such, may be more difficult to predict.

While gene family recruitment may be predictable if constrained by genetic solutions, the number and identity of genes encoding the function and whether they functionally specialize may be contingent on gene duplication mutations and cellular physiology, and therefore more difficult to predict. Our findings suggest that, while homologous sulfotransferase genes are recruited in the evolution of luciferin sulfation, the number and identity of luciferin sulfotransferase genes and their functional specialization may differ across convergent systems. Out of the luciferin sulfotransferases identified by at least one assay, four — ST2, ST3, ST4, and ST5 — can also sulfate p-nitrophenol. Conversely, in fireflies, Fallon et al. (2016) demonstrated only one of the three highly expressed sulfotransferase genes from the light-producing organ was capable of sulfating firefly luciferin and could not sulfate p-nitrophenol. Two other sulfotransferases readily sulfated p-nitrophenol but not firefly luciferin (Fallon et al. 2016). In sea pansies, Tzertzinis et al. (2022) tested multiple sulfotransferase isoforms, two of which can function as luciferin sulfotransferases (Tzertzinis et al. 2022). Therefore, although

fireflies might use a functionally specialized luciferin sulfotransferase to sulfate luciferin, ostracods may use many promiscuous luciferin sulfotransferases, and sea pansies may use one or both luciferin sulfotransferase isoforms. To test these hypotheses, further work, such as gene knockout experiments and gene co-expression analyses, would be necessary for confirming which sulfotransferases are associated with bioluminescence function *in vivo*.

Understanding how complex systems convergently evolved across levels of biological organization may inform general rules of predictability in evolution. The convergent bioluminescence systems of fireflies, sea pansies, and ostracods illustrate how a shared organismal function — using bioluminescence to deter predation — uses structurally different organs and tissues, and a mixture of non-homologous and homologous gene families to produce and modulate bioluminescence. Therefore, gene family recruitment in the evolution of bioluminescence may sometimes be more predictable than structural evolution, depending on the range of genetic solutions to a function. These findings may be broadly extended to the general rule that gene family recruitment may be more deterministic — and therefore predictable — if functions are limited to one or a few genetic solutions, and may be far less predictable if redundancy permits many non-homologous genetic solutions. However, even if the same gene family is recruited repeatedly, the number of genes encoding a biological function may vary across convergently evolved systems. While gene family recruitment may be predictable when functional evolution is genetically constrained, the number of genes encoding a

biological function — which may be contingent on gene duplication mutations and the diversity of mechanisms that fix them — is more difficult to predict.

Materials and Methods

We purchased TRIzol™ reagent (Catalog # 15596026), Terrific Broth (Catalog # BP9728500), and the pCMV-Cypridina Luc vector (Catalog # PI16150) — encoding ostracod luciferase — from Thermo Fisher Scientific (Waltham, MA). We purchased the pQE80L *E. coli* expression vector and Ni-NTA agarose (Catalog # R90115) from Qiagen (Hilden, Germany). The pET-6xHis-SUMO-MarathonRT plasmid (source of SUMO tag) was a gift from Dr. Anna Pyle (Addgene plasmid # 109029). We purchased *Cypridina* (ostracod) luciferin (purity > 99%, Catalog # 305-500) from Prolume LTD (Pinetop, AZ) and prepared single-use aliquots by dissolving 0.5 mg luciferin in 100 % methanol to a stock concentration of 6.15 mM, then stored the aliquots at -80 °C. To perform the luciferase assay, we obtained 3′-phosphoadenosine 5′-phosphosulfate (PAPS) lithium salt (purity ≥ 60 %, Catalog # sc-210759) from Santa Cruz Biotechnology (Dallas, TX), dissolved it in phosphate buffered saline (pH 7.4) to prepare 2 mM aliquots, and stored aliquots at -80 °C. To perform the sulfotransferase assay, we purchased the Universal Sulfotransferase Assay Kit (Catalog # EA003) and the lithium-free PAPS substrate (purity > 90%, Catalog # ES019) recommended for this assay from R&D Systems (Minneapolis, MN). The Universal Sulfotransferase Assay kit contained: (1) 0.5 µg/µL coupling phosphatase 3 in 20 mM Tris, 120 mM NaCl, 20 % glycerol, pH 7.5; (2) 5 mM PAP prepared in 10 mM sodium borate, pH 9.0; (3) 10X phosphatase buffer 3, comprised of 500 mM Tris, 150 mM $MgCl_2$, pH 7.5; (4) 1 mM phosphate standard

prepared in water; (5) Malachite green reagent A, comprised of ammonium molybdate in 3 M sulfuric acid; and (6) Malachite green reagents B, comprised of malachite green oxalate and polyvinyl alcohol. We obtained p-nitrophenol (purity ≥ 99%, Catalog # IC10246101) from MP Biomedicals (Santa Ana, CA), dissolved it in 20 mM tris hydrochloride (pH 7.4) to prepare a stock solution of 10 mM, and stored the stock at 4 °C. We purchased Chinese hamster ovary (CHO) cell lines from Takara Bio (San Jose, CA), oligonucleotide primers and gBlocks™ from Integrated DNA Technologies (Coralville, IA), and Q5 DNA polymerase (Catalog # M0491S), agarose gel extraction kit (Catalog # T1020S), reagents for Gibson assembly (Catalog # E2611L), competent *E. coli* 10-beta cells (Catalog # C3019H) and NEB Stable cells (Catalog # C3040H) (for plasmid propagation), and competent *E. coli* BL21 cells (Catalog # C2527H) (for protein expression) from New England Biolabs (Ipswich, MA). We purchased reagents for Quick Start™ Bradford protein assay (Catalog # 5000205) — including the Bovine Serum Albumin (BSA) standard for calibrating this assay (Catalog # 500-0201) — and Sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gels (Mini-PROTEAN TGX stain-free protein gels, Catalog # 4561094) from BioRad (Hercules, CA), Amicon® Ultra centrifugal filter units (30 kDa and 10 kDa molecular weight cutoff, Catalog # UFC903008 and UFC901008) from Sigma Aldrich (St. Louis, MO), and reagents for purifying plasmid DNA (Catalog # A1223) from Promega (Madison, WI). We purchased linear polyethyleneimine (25 kDa) transfection grade reagent (Catalog # 23966-100) from Polysciences (Warrington, PA), dissolved it in deionized water to prepare stock solutions at a concentration of 0.258 mg/mL, and stored stock solutions at -20 °C.

Construction of a high-quality reference transcriptome for Vargula tsujii

We used PacBio full-length isoform sequencing (Iso-Seq) technology to generate a high-quality reference transcriptome for the California luminous ostracod, *Vargula tsujii* (Supplemental Figure S3A). We collected and pooled total RNA from 88 *V. tsujii* specimens across various instar stages (carapace lengths, measured from the rostrum to the keel, ranging from 0.68 to 2.02 mm) (Supplemental Table S1), including adult males and brooding females (Supplemental Figure S2). We extracted RNA using TRIzol™ reagent, following the manufacturer's protocol (Invitrogen 2023) with two minor modifications. First, we used a higher centrifugation speed (16,100 x g) to pellet cell debris and separate the organic and aqueous layers. Second, we increased precipitation time to ~ 1 hour to increase RNA yield. We shipped purified RNA to the DNA Sequencing Center at Brigham Young University for library preparation and Single Molecule Real-Time (SMRT) sequencing (PacBio Sequel II). To maximize coverage of different sized transcripts, we used two fractions for sequencing: one that was resolved by pulsed field electrophoresis (BluePippin, Sage Science, Beverly, MA) to isolate 4-10 kb transcripts; and another that was not size-selected. We clustered and polished circular consensus sequencing (CCS) reads using the IsoSeq v3 pipeline (https://github.com/PacificBiosciences/IsoSeq). We combined the ensuing long reads with Illumina short reads from NCBI (SRR21201581) (Ellis et al. 2022) using the transcriptome assembler, rnaSPAdes, with the following flags: --rna, --only-assembler, and --trusted-contigs (Bushmanova et al. 2019). To reduce duplicated transcripts, we used cd-hit-est (Li and Godzik 2006; Fu et al. 2012) to cluster sequences with ≥ 95 %

identity. We assessed the completeness of the final assembly by using Benchmarking Universal Single-Copy Orthologs (BUSCO) v.3.0.2 (Simão et al. 2015; Waterhouse et al. 2018). BUSCO analysis identified 91.1 % of the total orthologs from the arthropod gene set, 80 % of which was classified as complete and 11.1 % fragmented.

Gene expression profiling and identification of sulfotransferase candidates in the upper lip of Vargula tsujii

To identify the most highly expressed sulfotransferase sequences in the light producing organ — known as the upper lip — of *Vargula tsujii*, we collected RNA and profiled gene expression (Supplemental Figure S3B) from the upper lips of ostracod specimens maintained in the Oakley lab culture (Goodheart et al. 2020). We extracted upper lips from nine ostracods — five males and four females — by dissection under a stereomicroscope, immediately mixed them with TRIzol™, and stored the samples at -80 °C until further use. For each individual ostracod, we extracted RNA — using the same extraction protocol as described above, for constructing the Iso-Seq transcriptome — and shipped purified RNA samples to the Genomic Sequencing and Analysis Facility at The University of Texas at Austin for individual library preparations and gene expression profiling using Tag-Seq (Illumina HiSeq 2500, SR100), resulting in an average of 1.1 million reads (range: 0.61-2.1 million) per sample after filtering (Supplemental Table 2) (Meyer et al. 2011). We processed Tag-Seq reads using a modified version of the Tag-Seq pipeline freely available on GitHub (https://github.com/z0on/tag-based_RNAseq) (Meyer et al. 2011; Lohman et al. 2016), which mapped the reads to a reference transcriptome — the Iso-Seq transcriptome of

V. tsujii — using Bowtie2 v.2.3.4.3 (Langmead and Salzberg 2012). On average, 0.87 million reads per sample (range: 0.45-1.5 million) produced unambiguous alignments to the reference transcriptome (Supplemental Table 2). We translated the reference transcriptome using TransDecoder v.5.5.0 (Haas, BJ. https://github.com/TransDecoder/TransDecoder), filtered for complete protein sequences, and used HMMER v.3.2.1 (hmmer.org) to identify protein sequences containing a sulfotransferase domain (Pfam ID: PF00685) with an e-value score less than or equal to 1e-3. We used MAFFT (v. 7.429) (Katoh and Standley 2013) to align protein sequences and used IQ-TREE (v.1.6.12) (Nguyen et al. 2015) to infer a maximum likelihood phylogeny, assuming the best-fit substitution model (LG+I+G4) — according to Bayesian Information Criterion — as determined with ModelFinder (Kalyaanamoorthy et al. 2017), and performed ultrafast bootstrap approximation (Hoang et al. 2017) with 1000 bootstrap replicates. To midpoint root the phylogeny and visualize bootstrap values and gene expression counts, we used the R packages phytools (v. 1.5-1) (Revell 2012) and ggtree (v. 3.9.0) (Yu et al. 2017).

Cloning, expression, and purification of ostracod (Cypridina noctiluca) luciferase

To facilitate protein purification, we used Gibson assembly (Supplemental Table S3) to incorporate 3 x FLAG and hexahistidine tags at the C-terminus of an ostracod (*Cypridina noctiluca*) luciferase encoding gene in the pCMV-Cypridina Luc vector (Supplemental Figure S5), which encodes for ampicillin resistance in bacterial cells and expresses the luciferase gene under a Cytomegalovirus (CMV) promoter. We propagated this modified vector by transforming chemically competent NEB Stable *E.*

coli cells, growing up a 5 mL overnight culture of transformed cells, and miniprepping the overnight culture to prepare plasmid DNA for sequencing. After DNA sequencing (Genewiz, South Plainfield, NJ) to confirm successful cloning, we co-transfected CHO cells with this vector and a marker plasmid encoding green fluorescent protein (GFP) by transient transfection of ~ 70% confluent cultures using polyethylenimine. We used fluorescence microscopy (Olympus CKX-41) to confirm transfection success by monitoring the production of GFP (Supplemental Figure S6A), then we harvested the cell culture media — containing secreted ostracod luciferase — 48 hours post transfection, and incubated the media with Ni-NTA agarose beads in 10 % (v/v) sample buffer (500 mM NaH_2PO_4, 1.5 M NaCl, 100 mM imidazole, pH 8) on a rotary shaker overnight at 4 °C. Subsequently, we loaded the Ni-NTA agarose slurry in a gravity-flow column, collected the flowthrough, and washed the column twice with 3 mL wash buffer (20 mM NaH_2PO_4, 500 mM NaCl, 10 mM imidazole, pH 7.4) to remove nonspecifically bound proteins. Then, we eluted the column twice with 2 mL elution buffer (20 mM NaH_2PO_4, 500 mM NaCl, 500 mM imidazole, pH 7.4). We assessed the protein content of each purification step by running SDS-PAGE and visualized protein bands by activating the stain free SDS-PAGE gel with UV (Supplementary Figure S6B). To confirm the presence of ostracod luciferase throughout purification, we added 20 μL of each purification step — namely, the flowthrough, wash, and each elution step — to 50 μL of 20 mM tris hydrochloride, pH 8.0. We then tested for luciferase activity by adding 5 μL of 50 μM luciferin to each sample and recorded luminescence using a microplate reader (Tecan Spark) with an integration time of 2 seconds. After confirming the presence of luciferase in our eluate samples (Supplemental Figure S6C), we

concentrated the eluate, exchanged it into a storage buffer (20 mM tris hydrochloride, 20 % glycerol, pH 8.0) by spin ultrafiltration using a centrifugal ultrafiltration unit with a 30 kDa molecular weight cutoff filter. We quantified protein yield using the Bradford assay (Bio-Rad Laboratories, Inc.), calibrated by using a standard curve of BSA. We flash-froze aliquots of recombinant luciferase in liquid nitrogen and stored them at -80 °C until use.

Characterizing ostracod (Cypridina noctiluca) luciferase kinetics

To characterize the enzyme kinetics of ostracod luciferase, we added 65 µL of purified ostracod luciferase to 10 µL of luciferin, resulting in a total volume of 75 µL and final concentrations 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 luminescence using a microplate reader (Tecan Spark) with an integration time of 2 seconds (N = 3). Since we observed decreasing light production at final substrate concentrations equal to or greater than 3.125 µM, we plotted and fitted these data using a Michaelis-Menten equation modified for modeling substrate inhibition (Copeland 2000) — $v = V_{max} [S] / K_m + [S] (1 + [S] / K_i)$ — where v = velocity, V_{max} = maximum velocity, S = substrate concentration, K_m = Michaelis-Menten constant, and K_i = the dissociation constant for the inactive ternary complex formed when another substrate molecule binds to an enzyme-substrate complex (Supplemental Figure S7).

Cloning, expression, and purification of Vargula tsujii sulfotransferase candidates

To express and purify candidate luciferin sulfotransferase proteins, we codon-optimized and synthesized DNA sequences corresponding to ST1-5 as gBlocks™ (Integrated DNA Technologies). We used Gibson assembly (Supplemental Table S3) to clone these fragments into a bacterial expression vector — pQE80L, containing a T5 promoter and encoding ampicillin resistance — allowing expression of the resulting proteins with an N-terminal hexahistidine tag for purification (Supplemental Figure S5). All plasmid constructs were propagated by electroporation of competent *E. coli* NEB® 10-beta cells, verified by DNA sequencing (Genewiz, South Plainfield, NJ), and subsequently electroporated in competent *E. coli* BL21 cells for protein expression. Initial efforts to express these proteins in *E. coli* BL21 resulted in soluble expression only for ST3 and ST4 (Supplemental Table S4, Supplemental Figure S8). For the remaining proteins — namely ST1, ST2, and ST5 — we attempted to improve expression by cloning the genes into a different expression vector — pET SUMO, containing a T7 promoter and encoding kanamycin resistance — which appended a small ubiquitin-like modifier (SUMO) tag at the N-terminus to improve protein solubility (Butt et al. 2005) (Supplemental Figure S5). Upon re-expression in *E. coli* BL21, the SUMO-tagged constructs showed clearly detectable bands at the expected size in a SDS-PAGE gel (Supplemental Table S4, Supplemental Figure S9).

We transformed *E. coli* BL21 cells with the sulfotransferase expressing plasmids and grew cells in 50 mL Terrific Broth with working concentrations of 100 µg/mL ampicillin or 50 µg/mL kanamycin at 37 ºC to mid-log phage (optical density at 600 nm ~ 0.4). Then, we induced protein expression by adding 50 µL isopropyl-β-D-thiogalactopyranoside

(IPTG) to a final concentration of 1 mM, moved induced cultures to an orbital shaker at room temperature, and continued protein expression for 18 hours. To harvest bacterial cells, we centrifuged the cell cultures (5000 x g, 10 minutes) 4 °C and froze cell pellets at -20 °C for 3 hours. We resuspended frozen cell pellets in lysis buffer (20 mM tris hydrochloride, 500 mM NaCl, 30 mM imidazole, 0.5 % tween 20, 5 mM β-mercaptoethanol, pH 7.4), lysed cells by sonication (5 cycles of 30 s on, 10 s off at 30 % amplitude, MiSonix Sonicator 4000 with microtip probe) on ice, pelleted cellular debris by centrifuging the crude lysate (12,000 x g, 30 minutes) at 4 °C, and incubated the clarified supernatant with Ni-NTA agarose beads at 4 °C overnight on a rotary shaker. Subsequently, we loaded the Ni-NTA slurry in a gravity-flow chromatography column and purified recombinant protein by washing twice with 2 mL wash buffer (20 mM tris hydrochloride, 500 mM NaCl, 30 mM imidazole pH 7.4), and eluting with 4 mL elution buffer (20 mM tris hydrochloride, 500 mM NaCl, 500 mM imidazole, pH 7.4). The eluates were concentrated and exchanged into a storage buffer (20 mM tris hydrochloride, 300 mM NaCl, 5 mM DTT, 20 % glycerol, pH 8) by spin ultrafiltration using centrifugal ultrafiltration units with 10 kDa molecular weight cutoff filters. We assessed protein purity by running SDS-PAGE, visualizing the bands by activating the stain-free gels with UV (BioRad Gel Doc), and performing densitometry analysis of protein bands on a SDS-PAGE gel using ImageJ (Supplemental Figure S9, Supplemental Table S5). We quantified the yield of purified protein by performing a Bradford assay (Bio-Rad Laboratories, Inc.), which was calibrated by using a standard curve of BSA. Using this approach, we successfully purified all five candidate proteins, achieving > 90 % homogeneity for the purified fractions, as determined by densitometry

analysis of protein bands resolved by SDS-PAGE (Supplemental Table S5). We freshly prepared and purified all sulfotransferase candidates for each experiment.

Assaying for sulfotransferase activity based on PAP formation

We tested sulfotransferase candidates by using the Universal Sulfotransferase Assay Kit (Catalog #EA003) (R&D Systems 2019), which measures the formation of 3′-phosphoadenosine-5′-phosphate (PAP) released as a byproduct during biochemical sulfation. In this assay, a coupling phosphatase 3 is used to hydrolyze PAP, releasing phosphate, which is quantified by adding malachite green and monitoring the increase in absorbance at 620 nm due to the formation of a complex between malachite green and phosphate (Prather et al. 2012). As recommended by the kit (R&D Systems 2019), we prepared all reagents in 50 mM Tris, 15 mM magnesium chloride, pH 7.5, hereafter referred to as “assay buffer”. Since this assay is sensitive to phosphate and lithium, we used phosphate and lithium-free PAPS substrate, supplied by R&D Systems as a 6.64 mM stock solution in 25 mM Tris, 23 % ethanol, pH 7.5. We set up the assay by incubating 5 μL coupling phosphatase 3 (100 ng/μL) with 10 μL ostracod luciferin (500 μM), 10 μL PAPS (1 mM), and 25 μL sulfotransferase (20 μM), resulting in a final volume of 50 μL and working concentrations of 10 ng/μL coupling phosphatase, 100 μM ostracod luciferin, 200 μM PAPS, and 10 μM sulfotransferase. We included a positive control to confirm the enzyme activity of the coupling phosphatase 3 by incubating 20 μL assay buffer with 5 μL coupling phosphatase (100 ng/μL) and 25 μL PAP (100 μM) in a total volume of 50 μL with working concentrations of 10 ng/μL coupling phosphatase and 50 μM PAP. Importantly, since coupling phosphatase exhibits activity with PAPS

and PAP at about a ratio of 1:350, we ran a negative control to measure this background activity by incubating 5 µL coupling phosphatase 3 (100 ng/µL) with 10 µL ostracod luciferin (500 µM), 10 µL PAPS (1 mM), and 25 µL of assay buffer, resulting in a final volume of 50 µL and working concentrations of 10 ng/µL coupling phosphatase 3, 100 µM ostracod luciferin, and 200 µM PAPS. Lastly, we included an assay blank, which contained 50 µL of the assay buffer, to quantify the background absorbance from the assay buffer. We incubated the reactions at room temperature for 3 hours, then measured the formation of phosphate from PAP — catalyzed by the coupling phosphatase — by adding malachite green and measuring the increase in absorbance at 620 nm with a microplate reader (non-normalized pathlength, Tecan Spark). We averaged the "blank" absorbance measurements and subtracted this average value from each of the absorbance measurements for the positive, negative, and experimental groups. To test whether the sulfotransferase candidates exhibited significant sulfation activity using the luciferin substrate, we used Dunnett's test to compare the absorbance measurements from each of the five experimental groups with the absorbance measurements from the negative control (N = 6).

In vitro modulation of bioluminescence by luciferin sulfation

We incubated sulfotransferase candidates with ostracod luciferin and PAPS, then measured their ability to enzymatically modulate bioluminescence by incubating the reaction product with ostracod luciferase and recording light production. To set up this experiment, we mixed 10 µL of 10 µM ostracod luciferin and 100 uM PAPS prepared in 20 mM tris hydrochloride pH 7.4, with 1 µL of each purified protein (final concentrations

ST1: 15.16 μM, ST2: 21.79 μM, ST3: 51.22 μM, ST4: 31.59 μM, ST5: 24.63 μM). We set up control reactions — as described above — but omitted PAPS to prevent sulfation activity. We incubated the reactions at 25 °C for 3 hours and terminated the reaction by heating at 65 °C for 5 minutes to denature the sulfotransferase candidates. To measure the decrease in bioluminescence due to the conversion of luciferin to a sulfated form, we cooled and added 90 μL of 20 mM tris hydrochloride pH 7.4 to dilute the reaction products. We then manually added 10 μL of the diluted reaction product to 55 μL purified ostracod luciferase (final concentration ~ 136 nM) in a microplate well, pipetted the mixture up and down once to ensure mixing, then immediately measured luminescence using a microplate reader (Tecan Spark) with an integration time of 2 seconds. To test whether each candidate sulfotransferase exhibited significant sulfation activity with 10 μM of luciferin substrate and 100 μM PAPS, as compared with the control reaction containing only 10 μM luciferin, we performed Student's t-tests (N = 3). To conclude whether *V. tsujii* contains a sulfotransferase gene with significant sulfation activity towards luciferin, we corrected for multiple comparisons by performing the Benjamini-Hochberg procedure to control the false discovery rate.

In vitro p-nitrophenol sulfation assay

We tested sulfotransferase candidates for their ability to sulfate p-nitrophenol, a common substrate for cytosolic sulfotransferases. To set up this assay, we mixed 25 μL of 16 μM purified luciferin sulfotransferase, 50 μL of 200 μM p-nitrophenol, and 25 μL of 2 mM PAPS resulting in a total reaction volume of 100 μL and working concentrations of 4 μM purified sulfotransferase, 100 μM p-nitrophenol, and 500 μM PAPS. We set up

control reactions — as described above — but omitted PAPS to prevent sulfation. We incubated the reactions at room temperature, and after 15 hours, tested for p-nitrophenol sulfation by measuring the decrease in absorbance at 405 nm with a microplate reader (non-normalized pathlength, Tecan Spark). We used Student's t-test to test whether each candidate sulfotransferase exhibited significant sulfation activity with the p-nitrophenol substrate and PAPS, as compared with the control reaction lacking PAPS (N = 3).

Phylogenetic analysis of functionally tested sulfotransferases

Based on literature searches (as of March 2023), we compiled sequence data from functionally tested sulfotransferases (Supplemental Table S6, Supplemental References), and used MAFFT (v. 7.429) (Katoh and Standley 2013) to align protein sequences. We inferred a maximum likelihood phylogeny using IQ-TREE (v. 1.6.12) (Nguyen et al. 2015), assuming the best-fit substitution model (LG+R5) — according to Bayesian Information Criterion — as determined with ModelFinder (Kalyaanamoorthy et al. 2017), and performed ultrafast bootstrap approximation (Hoang et al. 2017) with 1000 bootstrap replicates. To midpoint root the phylogeny and visualize bootstrap values, we used the R packages phytools (v. 1.5-1) (Revell 2012) and ggtree (v. 3.9.0) (Yu et al. 2017).

Acknowledgements

We thank members of the Oakley Lab, especially RM Varney and NM Hensley, and CM Pomory for discussion and feedback. This work was supported by National Science

Foundation (NSF) awards (IOS-1754770 and DEB-1457754 to THO), by 2019 (THO) and 2021 (AM) UCSB Academic Senate Faculty Research Grants, a R35 Maximizing Investigators' Research Award from the National Institutes of Health (5R35GM133530 to AM), a Discovery Award from the Department of Defense Congressionally Directed Medical Research Program (PR191154 to AM), and by the US Army Research Office under Cooperative Agreement Numbers W911NF-19-D-0001-0009 and ICB-2019-SSB-12 for the Institute for Collaborative Biotechnologies. ESL was supported by the NSF Graduate Research Fellowship Program (NSF GRFP 1650114) and JAG by the NSF Postdoctoral Fellowship in Biology (BIO PRFB-1711201). The authors acknowledge the use of the Biological Nanostructures Laboratory within the California NanoSystems Institute, supported by the University of California, Santa Barbara and the University of California, Office of the President. Use was made of 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.

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).

References

Atkinson DE. 1966. Regulation of Enzyme Activity. *Annu. Rev. Biochem.* 35:85–124.

Berens AJ, Hunt JH, Toth AL. 2015. Comparative transcriptomics of convergent evolution: different genes but conserved pathways underlie caste phenotypes across lineages of eusocial insects. *Mol. Biol. Evol.* 32:690–703.

Bessho-Uehara M, Huang W, Patry WL, Browne WE, Weng J-K, Haddock SHD. 2020. Evidence for de novo Biosynthesis of the Luminous Substrate Coelenterazine in Ctenophores. *iScience* 23:101859.

Bessho-Uehara M, Yamamoto N, Shigenobu S, Mori H, Kuwata K, Oba Y. 2020. Kleptoprotein bioluminescence: Parapriacanthus fish obtain luciferase from ostracod prey. *Sci Adv* 6:eaax4942.

Bio-Rad Laboratories, Inc. Quick Start Bradford Protein Assay. Available from: https://www.bio-rad.com/webroot/web/pdf/lsr/literature/4110065A.pdf

Blount ZD, Lenski RE, Losos JB. 2018. Contingency and determinism in evolution: Replaying life's tape. *Science* [Internet] 362. Available from: http://dx.doi.org/10.1126/science.aam5979

Bushmanova E, Antipov D, Lapidus A, Prjibelski AD. 2019. rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. *Gigascience* [Internet] 8. Available from: http://dx.doi.org/10.1093/gigascience/giz100

Chapman E, Best MD, Hanson SR, Wong C-H. 2004. Sulfotransferases: structure, mechanism, biological activity, inhibition, and synthetic utility. *Angew. Chem. Int. Ed Engl.* 43:3526–3548.

Christin PA, Weinreich DM, Besnard G. 2010. Causes and evolutionary significance of genetic convergence. *Trends Genet.* 26:400–405.

Clarke GL, Conover RJ, David CN, Nicol JAC. 1962. Comparative studies of luminescence in copepods and other pelagic marine animals*. *J. Mar. Biol. Assoc. U. K.* 42:541–564.

Cormier MJ, Kazuo H. 1964. Studies on the bioluminescence of Renilla renoformis. *Biochimica et Biophysica Acta (BBA) - Specialized Section on Biophysical Subjects* 88:99–104.

Coughtrie MWH. 2002. Sulfation through the looking glass—recent advances in sulfotransferase research for the curious. *Pharmacogenomics J.* 2:297–308.

Denton EJ, Herring PJ, Widder EA, Latz MF, Case JF. 1985. The roles of filters in the photophores of oceanic animals and their relation to vision in the oceanic

environment. *Proceedings of the Royal Society of London. Series B. Biological Sciences* 225:63–97.

Dilly PN, Herring PJ. 1978. The light organ and ink sac of Heteroteuthis dispar (Mollusca: Cephalopoda). *J. Zool.* 186:47–59.

Driscoll WJ, Komatsu K, Strott CA. 1995. Proposed active site domain in estrogen sulfotransferase as determined by mutational analysis. *Proceedings of the National Academy of Sciences* 92:12328–12332.

Duchatelet L, Delroisse J, Pinte N, Sato K, Ho H-C, Mallefet J. 2020. Adrenocorticotropic Hormone and Cyclic Adenosine Monophosphate are Involved in the Control of Shark Bioluminescence. *Photochem. Photobiol.* 96:37–45.

Ellis EA, Goodheart JA, Hensley NM, González VL, Reda NJ, Rivers TJ, Morin JG, Torres E, Gerrish GA, Oakley TH. 2022. Sexual Signals Persist over Deep Time: Ancient Co-option of Bioluminescence for Courtship Displays in Cypridinid Ostracods. *Syst. Biol.*:syac057.

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

Falaschi RL, Amaral DT, Santos I, Domingos AHR, Johnson GA, Martins AGS, Viroomal IB, Pompéia SL, Mirza JD, Oliveira AG, et al. 2019. Neoceroplatus betaryiensis nov. sp. (Diptera: Keroplatidae) is the first record of a bioluminescent fungus-gnat in South America. *Sci. Rep.* 9:1–9.

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.

Foster CSP, Van Dyke JU, Thompson MB, Smith NMA, Simpfendorfer CA, Murphy CR, Whittington CM. 2022. Different Genes are Recruited During Convergent Evolution of Pregnancy and the Placenta. *Mol. Biol. Evol.* 39:msac077.

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.

Goodheart JA, Minsky G, Brynjegard-Bialik MN, Drummond MS, David Munoz J, Fallon TR, Schultz DT, Weng J-K, Torres E, Oakley TH. 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.

Haddock SHD, Moline MA, Case JF. 2010. Bioluminescence in the Sea. *Ann. Rev. Mar. Sci.* 2:443–493.

Haddock SHD, Rivers TJ, Robison BH. 2001. Can coelenterates make coelenterazine? Dietary requirement for luciferin in cnidarian bioluminescence. *Proceedings of the National Academy of Sciences* 98:11148–11151.

Harvey EN. 1952. Bioluminescence Academic Press. *New York, NYUSA* 649.

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

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.

Hulet WH, Musil G. 1968. Intracellular Bacteria in the Light Organ of the Deep Sea Angler Fish, Melanocetus murrayi. *Copeia* 1968:506–512.

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.

Inouye S. 2007. Expression, purification and characterization of calcium-triggered luciferin-binding protein of Renilla reniformis. *Protein Expr. Purif.* 52:66–73.

Invitrogen. 2023. Invitrogen™ TRIzol™ Reagent Protocol. Available from: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/trizol_reagent.pdf

James MO. 2021. Enzyme Kinetics of PAPS-Sulfotransferase. *Methods Mol. Biol.* 2342:285–300.

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.

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

Kato S-I, Oba Y, Ojika M, Inouye S. 2004. Identification of the biosynthetic units of Cypridina luciferin in Cypridina (Vargula) hilgendorfii by LC/ESI-TOF-MS. *Tetrahedron* 60:11427–11434.

Kotlobay AA, Sarkisyan KS, Mokrushina YA, Marcet-Houben M, Serebrovskaya EO, Markina NM, Somermeyer LG, Gorokhovatsky AY, Vvedensky A, Purtov KV, et al. 2018. Genetically encodable bioluminescent system from fungi. *Proc. Natl. Acad. Sci. U. S. A.* 115:12728–12732.

Krönström J, Mallefet J. 2010. Evidence for a widespread involvement of NO in control of photogenesis in bioluminescent fish. *Acta Zool.* 91:474–483.

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

Lau ES, Oakley TH. 2020. Multi-level convergence of complex traits and the evolution of bioluminescence. *Biol. Rev. Camb. Philos. Soc.* [Internet]. Available from: https://onlinelibrary.wiley.com/doi/10.1111/brv.12672

Lee DH, Mittag M, Sczekan S, Morse D, Hastings JW. 1993. Molecular cloning and genomic organization of a gene for luciferin-binding protein from the dinoflagellate Gonyaulax polyedra. *J. Biol. Chem.* 268:8842–8850.

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.

Lohman BK, Weber JN, Bolnick DI. 2016. Evaluation of TagSeq, a reliable low-cost alternative for RNAseq. *Mol. Ecol. Resour.* 16:1315–1321.

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.

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.

Morin JG. 1983. Coastal Bioluminescence: Patterns and Functions. *Bull. Mar. Sci.* 33:787–817.

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.

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.

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.

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.

Orgogozo V. 2015. Replaying the tape of life in the twenty-first century. *Interface Focus* 5:20150057.

Pankey MS, Minin VN, Imholte GC, Suchard MA, Oakley TH. 2014. Predictable transcriptome evolution in the convergent and complex bioluminescent organs of squid. *Proc. Natl. Acad. Sci. U. S. A.* 111:E4736–E4742.

Peterson MK, Buck J. 1968. Light organ fine structure in certain Asiatic fireflies. *Biol. Bull.* 135:335–348.

Powell GS, Saxton NA, Pacheco YM, Stanger-Hall KF, Martin GJ, Kusy D, Felipe Lima Da Silveira L, Bocak L, Branham MA, Bybee SM. 2022. Beetle bioluminescence outshines extant aerial predators. *Proceedings of the Royal Society B: Biological Sciences* 289:20220821.

Prather B, Ethen CM, Machacek M, Wu ZL. 2012. Golgi-resident PAP-specific 3'-phosphatase-coupled sulfotransferase assays. *Anal. Biochem.* 423:86–92.

R&D Systems. 2019. R&D Systems Universal Sulfotransferase Activity Kit. Available from: https://resources.rndsystems.com/pdfs/datasheets/ea003.pdf

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

Rosenblum EB, Parent CE, Brandt EE. 2014. The Molecular Basis of Phenotypic Convergence. *Annu. Rev. Ecol. Evol. Syst.* 45:203–226.

Shimomura O. 2019. Bioluminescence: Chemical Principles and Methods. World Scientific

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212.

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.

Stayton CT. 2015. What does convergent evolution mean? The interpretation of convergence and its implications in the search for limits to evolution. *Interface Focus* 5:20150039.

Takenaka Y, Yamaguchi A, Tsuruoka N, Torimura M, Gojobori T, Shigeri Y. 2012. Evolution of bioluminescence in marine planktonic copepods. *Mol. Biol. Evol.* 29:1669–1681.

Thacker CE, Roje DM. 2009. Phylogeny of cardinalfishes (Teleostei: Gobiiformes: Apogonidae) and the evolution of visceral bioluminescence. *Mol. Phylogenet. Evol.* 52:735–745.

Thompson EM, Toya Y, Nafpaktitis BG, Goto T, Tsuji FI. 1988. Induction of bioluminescence capability in the marine fish, Porichthys notatus, by Vargula (Crustacean) [14C]luciferin and unllabeled analogues. *J. Exp. Biol.* 51:39–51.

Tsarkova AS. 2021. Luciferins Under Construction: A Review of Known Biosynthetic Pathways. *Frontiers in Ecology and Evolution* [Internet] 9. Available from: https://www.frontiersin.org/articles/10.3389/fevo.2021.667829

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.

Ujvari B, Casewell NR, Sunagar K, Arbuckle K, Wüster W, Lo N, O'Meally D, Beckmann C, King GF, Deplazes E, et al. 2015. Widespread convergence in toxin resistance by predictable molecular evolution. *Proc. Natl. Acad. Sci. U. S. A.* 112:11911–11916.

Van Belleghem SM, Ruggieri AA, Concha C, Livraghi L, Hebberecht L, Rivera ES, Ogilvie JG, Hanly JJ, Warren IA, Planas S, et al. 2023. High level of novelty under the hood of convergent evolution. *Science* 379:1043–1049.

Viviani VR, Silva JR, Amaral DT, Bevilaqua VR, Abdalla FC, Branchini BR, Johnson CH. 2020. A new brilliantly blue-emitting luciferin-luciferase system from Orfelia fultoni and Keroplatinae (Diptera). *Sci. Rep.* 10:9608.

Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol. Biol. Evol.* 35:543–548.

Widder EA. 2010. Bioluminescence in the Ocean: Origins of Biological, Chemical, and Ecological Diversity. *Science* [Internet] 328:704–708. Available from: http://dx.doi.org/10.1126/science.1174269

Woodard SH, Fischman BJ, Venkat A, Hudson ME, Varala K, Cameron SA, Clark AG, Robinson GE. 2011. Genes involved in convergent evolution of eusociality in bees. *Proc. Natl. Acad. Sci. U. S. A.* 108:7472–7477.

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.

Figures

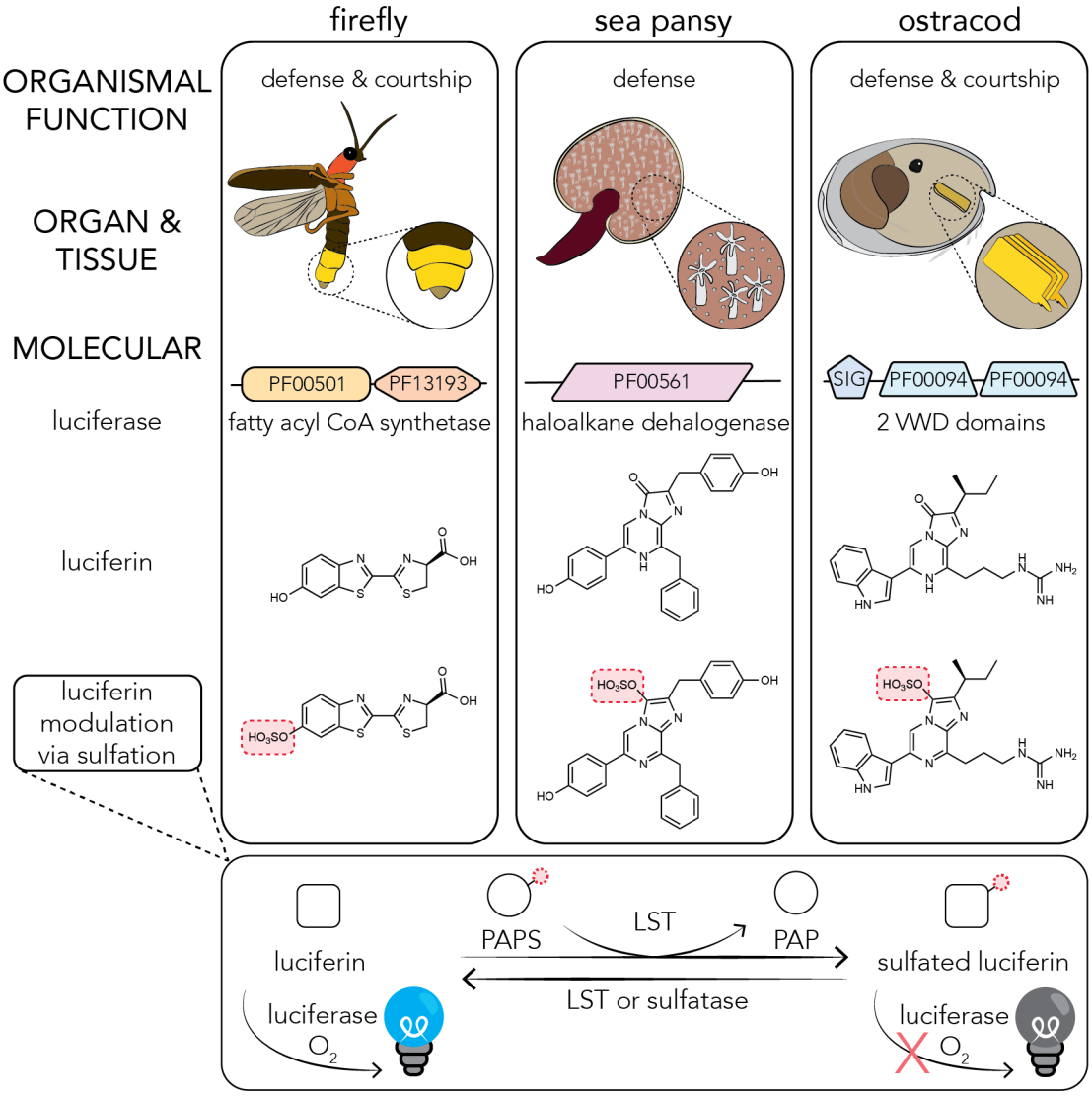

Figure 1. Convergently evolved bioluminescence systems of fireflies, sea pansies, and ostracods may vary in the extent of convergence across and within levels of biological organization. These taxa use bioluminescence as a defense mechanism but produce light by using structurally different organs and 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.

However, they use the same mechanism — sulfation — for modulating 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 is unable to react with their respective luciferase. The sulfate may be removed, either by a LST or sulfatase, to produce the active form of luciferin, which reacts with luciferase to produce light.

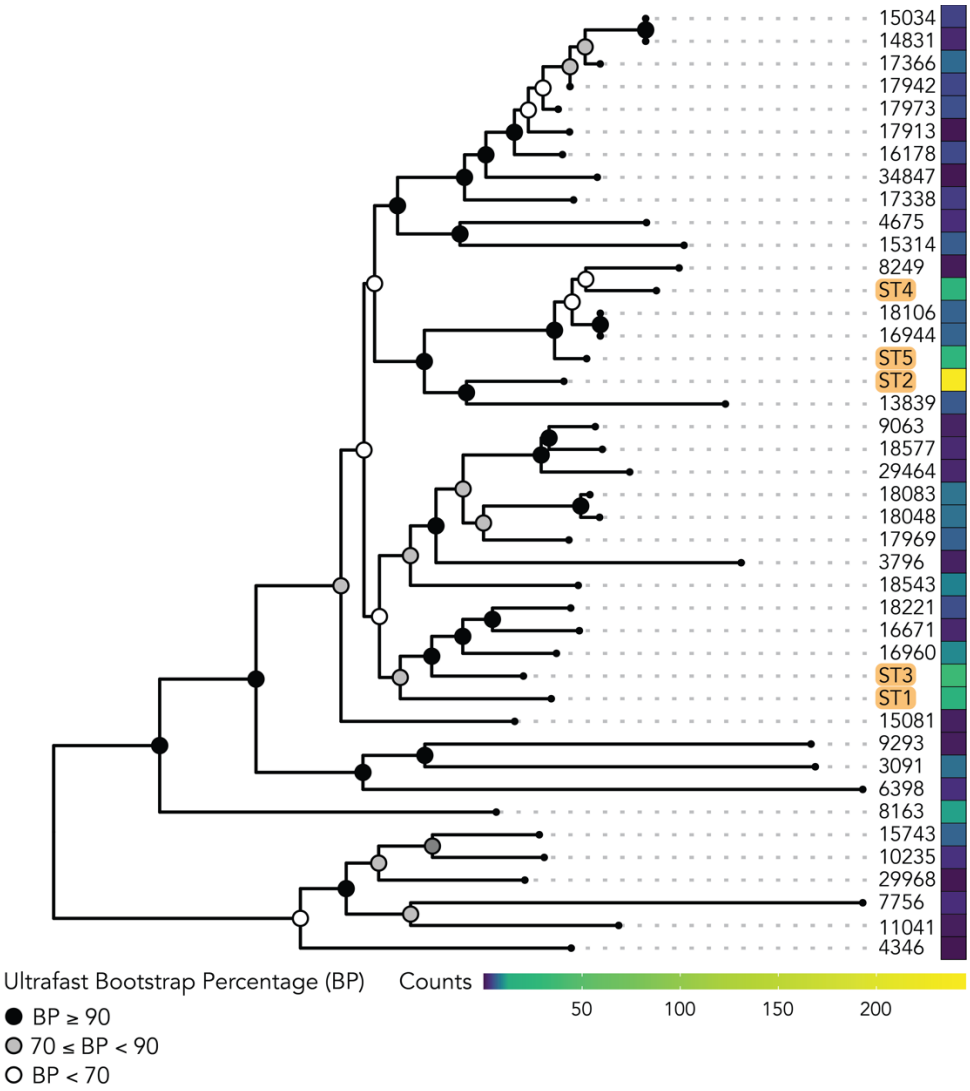

Figure 2. Maximum-likelihood inferred phylogeny and gene expression of sulfotransferase-domain containing proteins in *Vargula tsujii*. We inferred and mid-point rooted a maximum-likelihood phylogeny using candidate sulfotransferase sequences identified from the Iso-Seq reference transcriptome. The five most highly expressed candidates (labeled ST1-5) were selected for further biochemical testing. The color bar represents the average (N = 9 ostracods, Supplemental Figure S4) read counts in each ostracod sample. Ultrafast bootstrap values are represented by colored circles at each node.

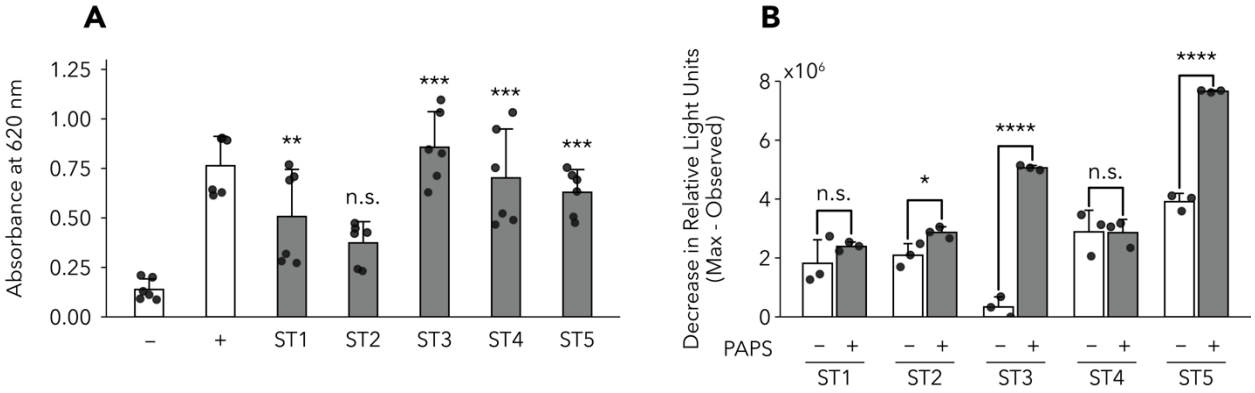

Figure 3. Multiple candidate sulfotransferases from *Vargula tsujii* exhibit luciferin sulfation activity. Sulfotransferase activity was assayed by incubating ostracod luciferin and PAPS with each of the 5 candidate enzymes for 3 hours. (A) The sulfotransferase assay assesses sulfotransferase activity by using a coupling phosphatase to hydrolyze PAP — a metabolic by-product of sulfation — to form phosphate. At the end of the reaction, malachite green is added to bind free phosphate, forming a complex that can be quantified via absorbance. Because coupling phosphatase is inhibited by lithium, we used lithium-free PAPS for this assay. Control reactions are shown in white bars and experimental groups are shown in gray bars. We included a positive (+) control — which consisted of 50 μL of 10 ng/μL coupling phosphatase and 50 μM PAP — to confirm the enzyme activity of the coupling phosphatase. The negative (–) control reaction contained 50 μL of 10 ng/μL coupling phosphatase, 100 μM luciferin, and 200 μM lithium-free PAPS, but lacked sulfotransferase. The experimental groups — ST1, ST2, ST3, ST4, and ST5 — contained 50 μL of 10 ng/μL coupling phosphatase, 100 μM luciferin, and 200 μM lithium-free PAPS, and 10 μM sulfotransferase. Compared to the negative control, four sulfotransferases — ST1, ST3, ST4, ST5 — showed significant luciferin sulfation activity. Error bars represent the standard deviation (N = 6). * denotes $P < 0.05$, **

denotes P < 0.01, *** denotes P < 0.001, and n.s. denotes non-significance (P ≥ 0.05). P values were calculated using Dunnett's test to compare the activity of the five STs against the negative control. (B) The luciferase assay uses luminescence to assess luciferin sulfation activity. If a candidate's luciferin sulfation activity is sufficient to modulate bioluminescence, we should observe a decrease in light production when ostracod luciferase is added to the product of their sulfation reaction. Experimental reactions, shown in gray bars, were set up by incubating 10 μL of 10 μM ostracod luciferin and 100 uM PAPS with 1 μL of each purified protein (final concentrations ST1: 15.16 μM, ST2: 21.79 μM, ST3: 51.22 μM, ST4: 31.59 μM, ST5: 24.63 μM). Control reactions, shown in white bars, were set up exactly as described, but PAPS was omitted to prevent sulfation activity. The high cause of variation between the negative controls may be due to the spontaneous oxidation of ostracod luciferin. Compared to the control reactions lacking PAPS, three candidates — ST2, ST3, and ST5 — exhibit a significant decrease in light production indicative of sulfation activity sufficient to modulate bioluminescence *in vitro*. Each datum is an average of three technical replicates. Error bars represent the standard deviation (N = 3). * denotes P < 0.05, ** denotes P < 0.01, *** denotes P < 0.001, **** denotes P < 0.0001, and n.s. denotes non-significance (P ≥ 0.05). P values were calculated by performing unpaired Student's t-tests between STs incubated with and without PAPS.

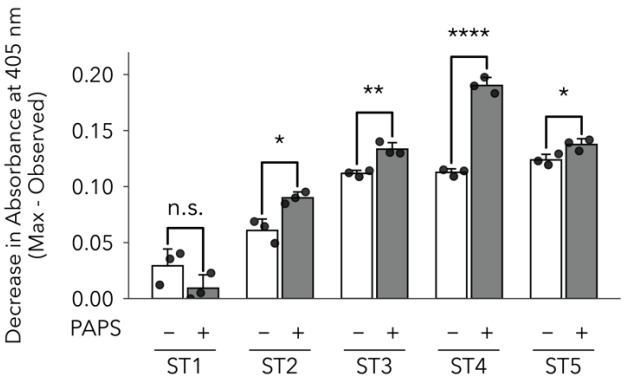

Figure 4. Four candidates sulfate p-nitrophenol. Sulfation of p-nitrophenol, a common substrate for many sulfotransferases, was assayed by measuring change in absorbance at 405 nm following a 15 hour incubation with purified sulfotransferase proteins in the absence and presence of PAPS. Control reactions, 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 significant p-nitrophenol sulfation under the conditions tested. Error bars represent the standard deviation (N = 3). * denotes $P < 0.05$, ** denotes $P < 0.01$, *** denotes $P < 0.001$, **** denotes $P < 0.0001$, and n.s. denotes non-significance ($P \geq 0.05$). P values were calculated by performing unpaired Student's t-tests between STs incubated with and without PAPS.

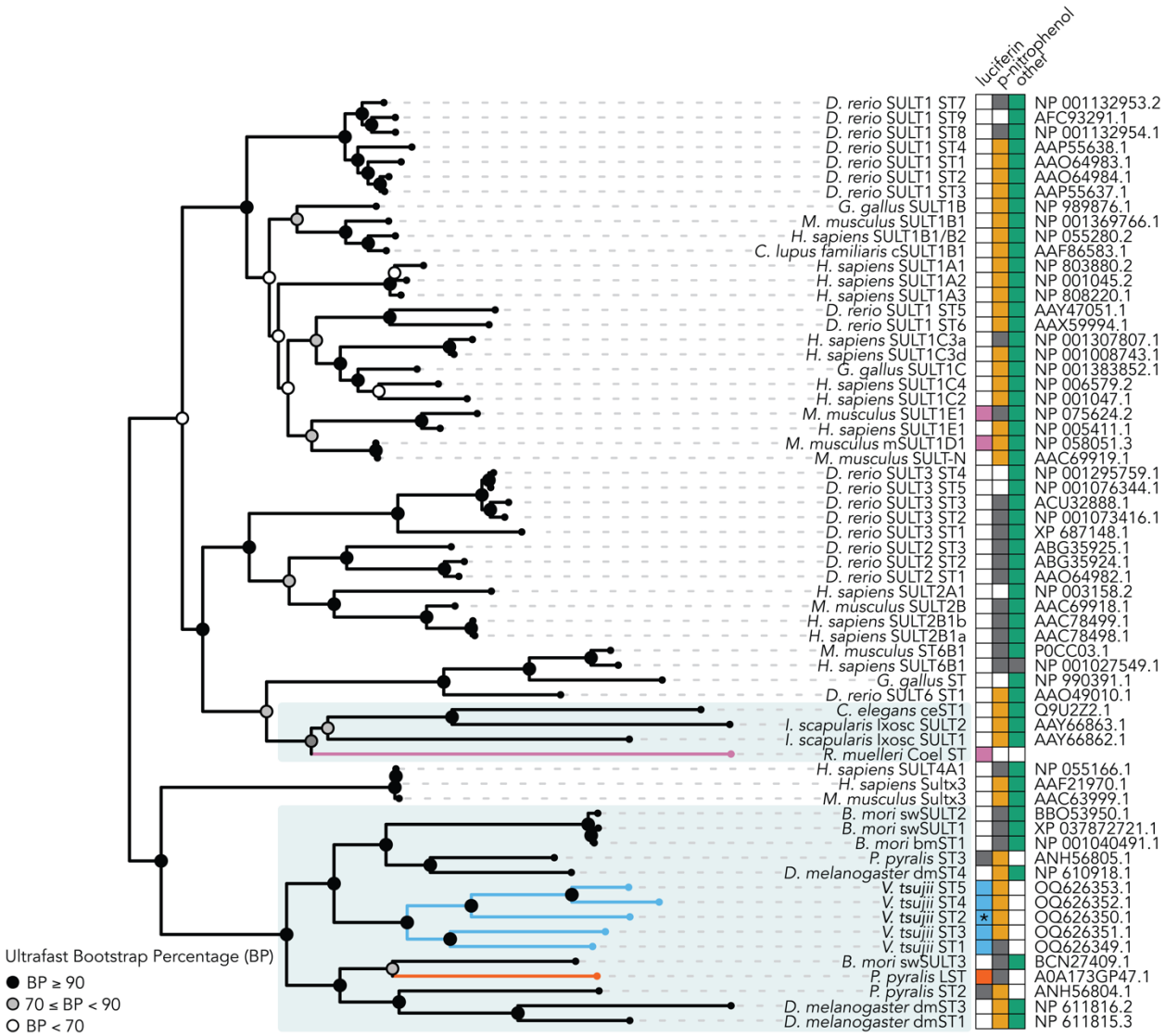

Figure 5. Maximum-likelihood inferred phylogeny of functionally tested sulfotransferases. Luciferin sulfotransferases from luminous organisms are denoted by colored branches (firefly = red, ostracods = blue, sea pansies = purple). Invertebrate sulfotransferases are highlighted in the shaded boxes. Colored squares (red = firefly luciferin, blue = ostracod luciferin, purple = sea pansy luciferin, orange = p-nitrophenol, green = other) indicate functionally demonstrated substrates for each sulfotransferase; gray squares indicate molecules that cannot be sulfated by the respective sulfotransferase; white squares indicate molecules that have not yet been tested.

Accession numbers for each sulfotransferase are listed on the right. Ultrafast bootstrap values are represented by colored circles at each node. * *V. tsujii* ST2 does not exhibit statistically significant luciferin sulfation activity after correcting for multiple comparisons