Molecular evolution of luciferase diversified bioluminescent signals in sea fireflies

Nicholai M. Hensley*, Emily A. Ellis, Nicole Y. Leung, John Coupart, Alexander Mikhailovsky, Daryl A. Taketa, Michael Tessler, David F. Gruber, Anthony W. De Tomaso, Yasuo Mitani, Trevor J. Rivers, Gretchen A. Gerrish, Elizabeth Torres, Todd H. Oakley*,**

* Equal contributors
Corresponding author:
**oakley@ucsb.edu

Classification: Biological Sciences - Evolution

Key words: diversification, genetic basis, drift, constraint, positive selection, bioluminescence, c-luciferase, Ostracoda

Abstract (209/250)
Understanding the genetic causes of evolutionary diversification is challenging because differences across species are complex, often involving many genes. However, cases where single or few genetic loci affect a feature that varies dramatically across a radiation of species would provide tractable opportunities to understand the genetics of diversification. Here, we show the diversification of bioluminescent signals in cypridinid ostracods (“sea fireflies”) to be strongly influenced by a single gene, cypridinid-luciferase. We find different evolutionary processes, including selection, drift, and constraint, each acted on c-luciferase at different times during evolutionary history and impacted different phenotypes, diversifying behavioral signals across species. In particular, some amino acid sites in c-luciferase evolved under episodic diversifying selection, and are associated significantly with phenotypic changes in both enzyme kinetics and color, which impact signals directly. We also find that multiple other amino acid positions in c-luciferase evolved neutrally or under purifying selection and may have impacted the variation of color of bioluminescent signals across genera. This work provides a rare glimpse into the genetic basis of diversification across many species, showing how multiple evolutionary processes may act at different times during a radiation of species to diversify phenotypes. These results indicate not only selection but also drift and constraint may be important evolutionary drivers of species diversification.

PNAS Significance statement (100/120 words)
A hallmark of life is its astounding diversity. While we are beginning to understand the drivers of biodiversity, uncovering the genetic basis remains challenging. As such, how different molecular evolutionary processes act to diversify phenotypes is a major question in biology. Here we show a single gene to be important in a riotous diversity of fantastical behaviors - the bioluminescent signals of sea fireflies - allowing us to demonstrate multiple evolutionary forces including

selection, drift, and constraint contributed to diversification. Our work highlights that not only selection but also neutral processes and constraint have each worked at different times to shape phenotypic diversity.

Main

Why and how some groups of species diversify more than others are enduring questions in biology, with broad implications for the origin and maintenance of biodiversity. Particularly challenging is to understand the genetic underpinnings of diversification, because numerous genes typically underlie quantitative phenotypic differences that vary across many species (1–3). In addition, wide disparities across species also exist in particular traits like morphology and behaviors including courtship that often accompany rapid speciation (4–8). During rapid divergence of particular traits, we might expect few loci of large effect (e.g. effector genes *sensu (9)*) or with a simple genetic architecture (10) may be better poised to diversify phenotypes quickly (11, 12), but such genes are often difficult to identify. Rapidly diversifying phenotypes have therefore inspired genome-scale studies of species radiations, which show that positive and purifying selection are involved in species differences and sometimes can act on a limited number of loci to promote such variety (13, 14). When possible, another approach to the genetics of diversification is to identify cases where one or a few genes are linked to diverse phenotypes across many species. Visual pigment genes (opsins) are a prime example of a gene family that underlies an organismal phenotype (color sensitivity) with a shared genetic basis across species (15–17). While most genomic and single-gene studies highlight how selection and/or epistasis impact phenotypic diversification (14, 18), a role for neutral processes has been difficult to demonstrate (19).

Bioluminescent ostracods (family Cypridinidae) and their phenotypically disparate signals provide a particularly attractive system for integrative understanding of diversification (5, 20–23). Bioluminescent cypridinids globally use light as an anti-predator display, including Caribbean species that also use light for courtship. This Caribbean clade diversified into dozens of species, each with a courtship display that is conserved within, but variable between species (22, 24). Although both sexes use light for anti-predator displays by expelling light-producing chemicals mixed with mucus, as far as we know, only males of Caribbean species produce courtship signals. In contrast to the light cloud produced during anti-predator signals, courtship signals are comprised of delicate, discrete pulses of light formed as males rapidly swim (23, 25). Multiple species commonly live in geographical sympatry, yet produce signals above different microhabitats. These courtship signals also vary in other parameters, including display angle (with reference to the benthos), specific time of onset (26), and time and distance between light pulses (21). Pulses can vary in brightness, kinetics, and color (20, 23, 27). Once ostracods externally secrete the pulses in bioluminescent signals, the phenotype of a single pulse is dictated by biochemistry instead of behavior, and depends largely on well-understood chemical reactions between cypridinid luciferase (“c-luciferase”) and the substrate, luciferin. Because the substrate is shared within this ostracod family (27), biochemical differences in light production largely depend on differences in c-luciferases. Therefore, cypridinid ostracods provide a system whereby some critical aspects of signals may be separated from the complexities of behavior

and connected directly to genetic changes in enzymes, allowing unique insights into the relationships between genotype, phenotype, and diversification.

The decay rate of light production is one target for connecting genotype, phenotype, and diversification of signals in cypridinid ostracods. Recent work shows decay rates vary considerably across species of luminous ostracods, and are correlated with the temporal duration of light in the individual pulses of courtship signals (20). Hensley et al. (2019) measured kinetics of light production by fitting an exponential decay parameter to each species' bioluminescence. Cypridinid bioluminescence is a single-order biochemical reaction whose rate depends on the concentration of substrate (28). Although other components within the mucus, including varying levels of luciferin, alter the overall reaction dynamics, light production fades exponentially once substrate becomes limiting and is a reliable metric of c-luciferase function. We hypothesize that differences in sequences of c-luciferase enzymes influence differences in the light decay kinetics, which vary dramatically across species and affect the light duration in both anti-predation and courtship displays. The duration of anti-predator light could be under natural selection, and the duration of courtship pulses could be under sexual selection, perhaps used for species recognition, and/or choosing amongst potential conspecific mates.

In addition to kinetics of light decay, the color of cypridinid bioluminescence could be dictated by differences in c-luciferase proteins. However unlike kinetics, emission spectra ("colors") are not well-characterized in many cypridinids. Previously published experiments hinted at variation in color of ostracod bioluminescence (27, 29), yet interspecific comparisons were impossible due to differences in methods and lack of replication (Table S1). Harvey (27) first noticed a difference in color between two species when he cross-reacted crude preparations from *V. hilgendorfii* and an unknown Jamaican species. He noted the luciferase preparation of *V. hilgendorfii* catalyzed a "bluish" light and the Jamaican luciferase a "yellowish" light, concluding that the protein (luciferase) dictated the color. The qualitative observations of Harvey nearly 100 years ago are consistent with the more recently published emission spectra of ostracods (29–32) that suggest some Caribbean species have higher λ_{max} values than other species. If so, λ_{max} may be another phenotype that varies across species and is dictated by c-luciferase. Like the rate of light decay, differences in color of bioluminescence could be functionally neutral, or could be influenced by natural selection, mediated through its appearance to would-be predators, and/or sexual selection, mediated through its appearance to would-be mates.

Because c-luciferases may be expressed *in vitro* to test biochemical functions (33) - including light production, decay kinetics, and color - this system fits squarely into the "functional synthesis" research paradigm (34), which combines phylogenetic analyses of sequences and manipulative biochemical experiments to allow mechanistic understanding of adaptation and evolutionary constraints. However, despite the potential value of bioluminescence for genotype-phenotype relationships (35, 36), previous studies that characterize luciferase genes of cypridinid bioluminescence exist for only two species (32, 37), which show cypridinid luciferases evolved independently of other luciferases, have a signal peptide leading to secretion outside cells, and possess two Von Willebrand Factor-D (VWD) domains (38). Here, we characterize new c-luciferase genes for phylogenetic sequence analyses in combination with functional characterizations of bioluminescence. In particular, we

hypothesize that differences in biochemical kinetics, namely the rate of decay of light production ("decay"), and emission spectra of bioluminescence ("color") can be linked to genetic variation in c-luciferase genes during the diversification of both anti-predator and courtship signals, to allow us to understand the molecular evolutionary processes that shaped

Results

Thirteen new putative luciferases. We identified 13 new putative luciferases from whole-body transcriptomes of cypridinid ostracods. These genes form part of a monophyletic family along with the only two previously published cypridinid luciferases (Fig. 1; Fig. S1). These putative luciferases bear multiple hallmarks of c-luciferase, including two Von Willebrand Factor-D (VWD) domains, a signal peptide, and conserved cysteine sites (32, 37, 38).

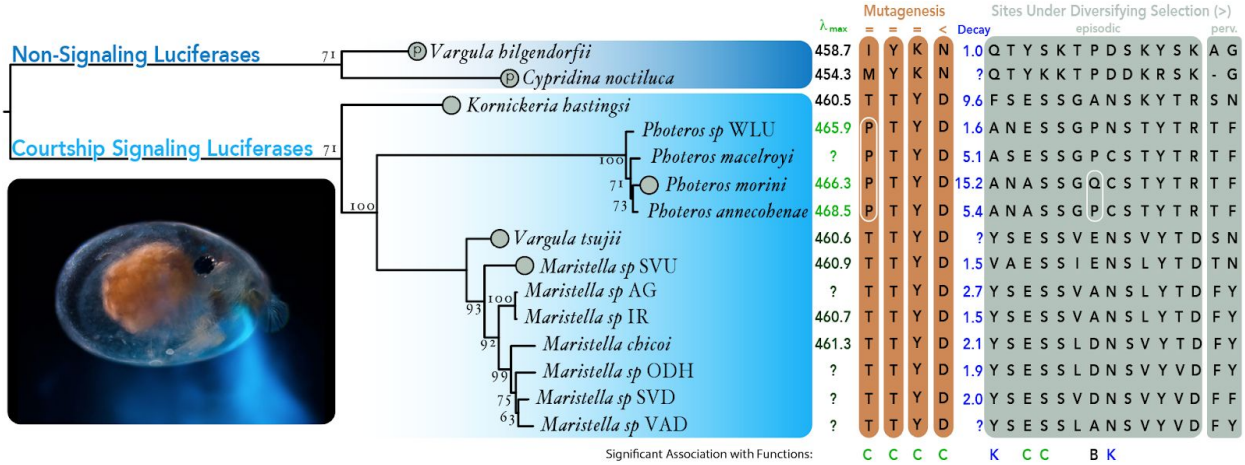


Figure 1. Maximum likelihood phylogeny of new and published c-luciferase genes using codon-aligned DNA and GTR model with fast bootstraps (percentages at nodes) in IQ-Tree. Circles at tips of phylogeny illustrate six genes expressed in vitro that show light catalysis. Two circles with "p" inside are previously published, the other four are explained herein (Fig 2). At right are two biochemical phenotypes (λ_{max} and light decay constant), and interesting amino acid sites, including three that affect λ_{max} in mutagenesis experiments (orange shading), thirteen under episodic diversifying selection using the MEME approach of hyphy and two under pervasive diversifying selection using the FEL approach of hyphy. Numbers of sites at bottom correspond to the unaligned *Cypridina noctiluca* luciferase sequence. We label the dn:ds ratio above the sites with ">" meaning diversifying/positive selection, "<" meaning purifying selection, and "=" meaning failure to reject null neutral model. White ellipses around sites are amino acid patterns discussed in the main text. We label sites significantly associated with Color (C), Kinetics (K), and Both color and kinetics (B). Table 1 has p-values for these sites. From left to right, site numbers are (λ_{max} values), 38, 178, 375, 404, (Decay values), 19, 67, 74, 114, 132, 148, 160, 232, 256, 262, 289, 358, 445, 20, 180.

Patterns of natural selection in c-luciferases. Using mixed effects maximum likelihood in HYPHY (39), we found ratios of synonymous to non-synonymous substitution rates in c-luciferases to indicate twelve sites consistent with episodic diversifying selection (Fig. 1), distributed throughout the gene. "Episodic" diversifying selection refers to sites with elevated dn:ds along a proportion of branches of the gene tree. According to Fixed Effects Likelihood (FEL) analysis (40), 250 c-luciferase sites have significantly low rates of dn:ds, consistent with purifying

(negative) selection, and two sites with significantly high dn:ds, consistent with pervasive positive selection, or sites with elevated dn:ds for the entire gene tree.

Luciferases from Exemplar Species are Functional. Using *in vitro* expression in mammalian and/or yeast cells (Fig 2), we tested the ability of putative c-luciferases to catalyze light reactions. We selected four exemplar species representing different genera of bioluminescent Cypridinidae. For each luciferase construct, light levels increased significantly after the addition of the substrate luciferin or compared to negative controls (Fig. 2; Table S2). Adding luciferin to biological media often produces light (41) that varies across biological replicates. We note this variation, yet also report statistically significant differences in light production after adding substrate (for yeast) or between c-luciferase secreting-cells and cells or media alone (for HEK293 cells and yeast) (Fig 2). This is consistent with the putative c-luciferases across multiple genera of Cypridinidae being functional c-luciferases.

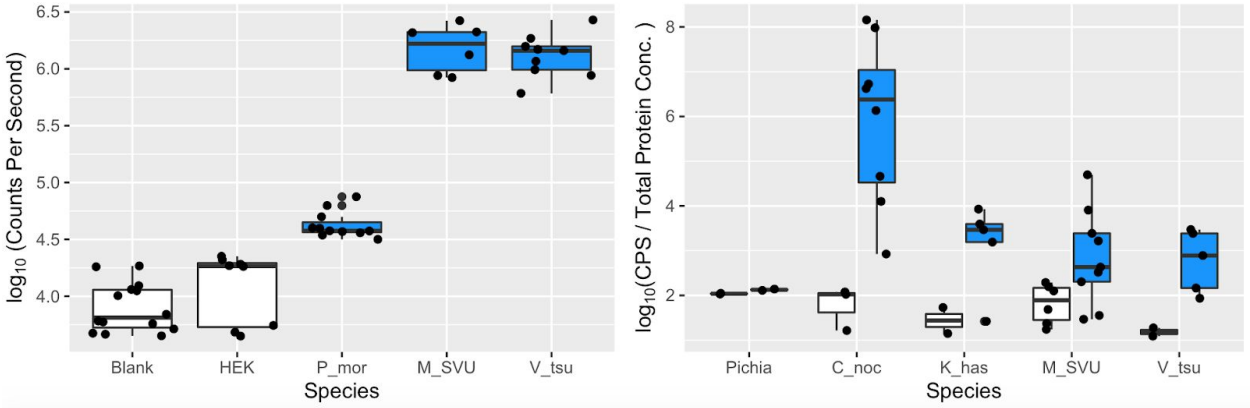


Figure 2. Exemplar luciferases are functional. **A**. Light produced after addition of luciferin to putative c-luciferase constructs from different species, measured in log Counts Per Second. Controls include blank cell culture media and HEK cells with no construct. Constructs are: (1) *Photeros morini* (P_mor), *Maristella sp SVU* (M_SVU), and *Vargula tsujii* (V_tsu) **B**. Log counts per second normalized by total protein concentration for four c-luciferase constructs expressed in yeast *Pichia*. Measures are before the addition of the substrate luciferin (grey) and after (blue). Control is *Pichia* cells alone. Constructs are: (1) the host strain of *Pichia* without a luciferase as a negative control, (2) a sequence known from *Cypridina noctiluca* (C_noc) as a positive control, (3) a novel sequence from *Kornickeria hastingsi carriebowae* (K_has), (4) a novel sequence from an undescribed species from Belize, "SVU" (nominal genus *Maristella*), and (5) a novel sequence from the California sea firefly *Vargula tsujii* . Each datum is an average of three technical replicates.

Emission spectra vary across species of bioluminescent Cypridinidae. By crushing whole specimens to elicit light production in front of a spectroradiometer (see Methods), we obtained new data on emission spectra from 20 species (Fig. 3; Tables S3, S4). The wavelength of maximum emission (λ_{max}) varies from 458.7 ± 1.80 nm in *V. hilgendorfii* to 468.0 ± 1.80 nm in *Photeros sp. EGD*. Previously published data extend this range, with *Cypridina noctiluca* at 454.3 nm and *Photeros graminicola* at 471.1 nm. We found λ_{max} from species of *Photeros* do not overlap in value with λ_{max} of other species. The lowest λ_{max} from any *Photeros* we measured is

465.6 ± 0.78 nm, whereas the highest of any non-Photeros species is 461.5 ± 2.70. Assuming *Photeros* is monophyletic, we infer an evolutionary increase in λ_{max} along the branch leading to *Photeros*. Monophyly of *Photeros* is supported by published morphological (42) and molecular (20) phylogenetic analyses, by the phylogeny of c-luciferases we present here (Fig. 1), and by a diagnostic ratio of length:height of carapace (21, 43, 44), which we also used here to classify undescribed species into genera (Table S5). Full Width Half Maximum (FWHM) is a common parameter to describe the variation (width) of wavelengths in emission spectra. We found FWHM values to also vary between species of Cypridinidae (Fig. S2), ranging from 75.05 ± 0.91 to 85.44 ± 0.32, although we notice no apparent phylogenetic pattern.

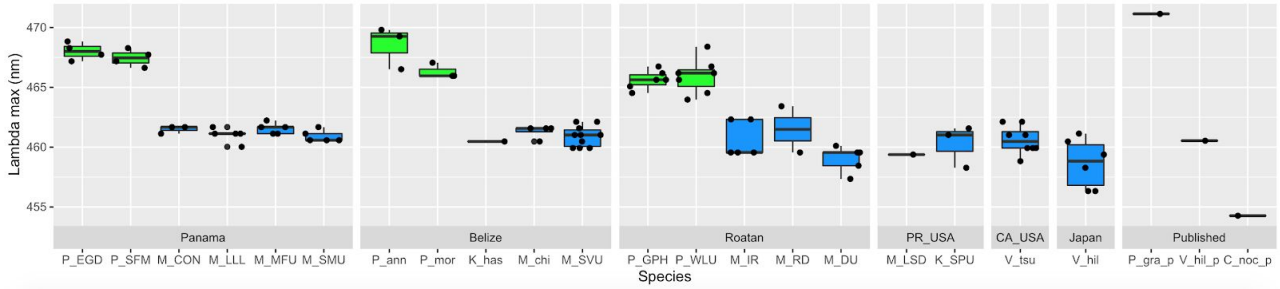


Figure 3.*Photeros* (green) have higher values of λ_{max} than other genera (blue). Wavelength value with highest emission (λ_{max}; "Lambda max" on the y-axis) in nanometers (nm) from new emission spectra from 20 species and previously published spectra from 3 species.

Genetic basis of phenotypic changes. By analyzing previously published mutagenesis studies along with new c-luciferase sequences and new color data, we identified amino acid sites that influence λ_{max}. The site most strongly associated with λ_{max} is 178 (ANOVA, p=2.2 e-16; Table 1). This effect is clear from published mutations alone: wild-type luciferase from *Cypridina noctiluca* (hereafter *Cn-luc*) had λ_{max} of 454 nm with a methionine at 178 (M-178-M) while mutants had lower λ_{max} (M-178-R at 435 nm; M-178-K at 447 nm), although these sequences also had other mutations at sites besides 178 (45). Site 178 also varies in c-luciferases of living species (Fig. 1), and notably all *Photeros*, the genus with green-shifted λ_{max}. In addition to site 178, variation at sites 280, 375, and 404 shows significant correlation with $\lambda_{max.}$ (Fig. 1; Table 1). Our ANOVA showed a significant interaction between sites 375 and 404, which may be interpreted as epistatic interactions (46).

Table 1. Sites significantly associated with phenotypes and their inferred processes of molecular evolution based on estimated rates of synonymous to non-synonymous substitution ratios.

Phenotype	Site (Cypridina)	Site (Aligned)	ANOVA p	Process	Hyphy p
Lambda-max (color)	178	207	2.2 e-16	Neutral	NS
	280	311	2.5 e-06	Neutral	NS
	375	406	0.013	Neutral	NS
	404	435	0.011	Purifying	0.007
	375:404	406:435	0.013	Epistasis	NA
	74	102	0.0001	Positive	0.0906
	114	142	0.0033	Positive	0.0015
	160	189	0.0053	Positive	0.0509
Light decay rate (kinetics)	19	41	0.0002	Positive	0.0256
	160	189	7.99E-05	Positive	0.0509
	232	261	0.003	Positive	0.0005

c-Luciferase and the diversification of function. Comparison of c-luciferases and bioluminescence phenotypes shows how multiple evolutionary forces acted on one gene, diversifying courtship signals. We first examined patterns of non-synonymous to synonymous (dn:ds) rates of mutation in the four sites from mutagenesis experiments that most strongly influence λ_{max}. Three of these sites (178, 280, 375) failed to reject the null model of neutral evolution. One site (404) had low dn:ds (p=0.071) in the FEL analysis of hyphy. Each of these sites changed λ_{max} in mutagenesis experiments and three of them (280, 375, 404) are fixed within courtship-signaling cypridinids, fixed within non-courtship species, but different between courtship-signaling and non-courtship cypridinids.

In addition to mutated sites, we checked sites under positive selection (dn>ds) for significant associations with biochemical functions. Of the thirteen sites under episodic diversifying selection, five are significantly correlated to bioluminescence phenotypes. Site 160 is under diversifying (positive) selection and correlated significantly with both λ_{max} and light decay of c-luciferases (Fig. 1; Table 1). Four additional sites under positive selection are also correlated with function: sites 74 and 114 are correlated with λ_{max} and sites 19 and 232 are correlated with rates of light decay. These results show how different processes of molecular evolution, including positive and purifying selection, and neutral evolution, together contributed to diversification of biochemical phenotypes that may be important for cypridinid signals (Fig. 4).

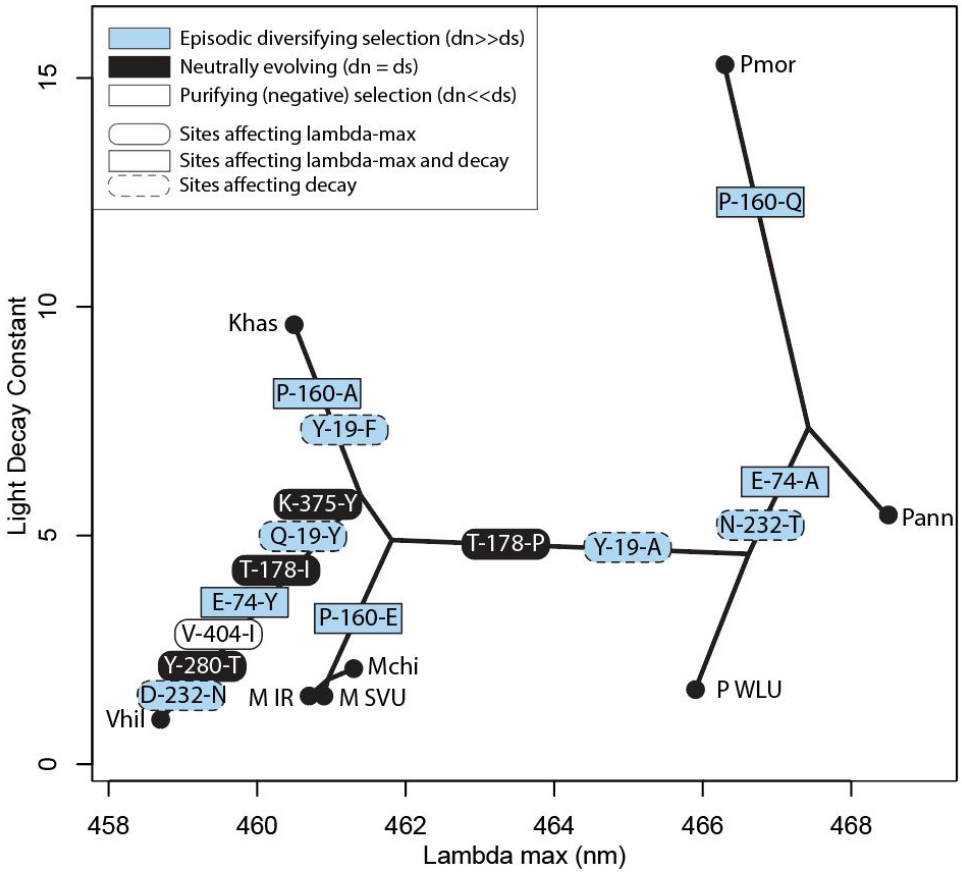


Figure 4. Bioluminescent phenotypes, including λ_{max} ("Lambda max" on the x-axis) of emission spectra and decay of light emission ("Light decay constant on the y-axis), diversified during the evolution of Cypridinidae, perhaps through different mutations in c-luciferases, and influenced by different evolutionary processes. Circles represent points that different species for which we have (N = 8) occupy in this phylophenospace of bioluminescence, connected by branches representing phylogenetic relationships. Shapes and colors on branches represent inferred mutations and evolutionary processes acting on c-luciferases and correlated with bioluminescence phenotypes. Rectangular sites are correlated with both decay and color, sites with rounded edges either affect color (solid border) or correlate with decay (dashed border). The color of the shapes represent evolutionary processes inferred at those sites, blue is positive selection (dn>>ds), white is purifying selection (dn<<ds), and black is neutral evolution. Each site has a number corresponding to the homologous site in *Cypridina noctiluca* and two single-letter abbreviations for amino acids representing evolved changes at that site. Species abbreviations are Vhil = *Vargula hilgendorfii,* Khas= *Kornickeria hastingsi*, M IR = *Maristella sp.* IR, M SVU = *M. sp.* SVU, Mchi = *M. chicoi*, P WLU = *Photeros sp* WLU, Pann = *P. annecohenae*, Pmor = *P. morini*.

Discussion

The genetic basis of species diversification remains understudied because many species differences involve many genetic differences across species. At the same time, scientists know very few cases where a single gene contributes to important and extensive phenotypic variation across radiations of species - which could provide tractable systems to study the molecular

evolution of diversification. The bioluminescent signals of cypridinid ostracods provide a prime example of signals that vary in multiple parameters across species, and here we show signal diversity to be impacted in important ways by at least a single gene, c-luciferase. Only two previously published c-luciferase genes existed in the literature and we report multiple new luciferase genes and biochemical phenotypes from diverse cypridinid species. Our results indicate various evolutionary forces acted on c-luciferase, contributing to diversification of phenotypes. In addition to some sites in c-luciferase that control color evolving neutrally or under purifying selection, we found episodic diversifying selection on luciferases at amino acid sites strongly associated with changes in both color and enzyme kinetics. In addition to connecting specific mutations to the diversification of biochemical functions, it is interesting to hypothesize how organismal and biochemical phenotypes relate to each other during evolution of this system. As with most biological phenomena, variation in phenotypes could be directly under selection, non-functional, and/or influenced by phylogenetic or biochemical constraints.

First, differences in molecular phenotypes could be shaped by natural and/or sexual selection. Patterns of variation in c-luciferase sequences support selection as one driver of phenotypic variation of biochemical properties of light production. We found multiple sites (19, 160, and 232) in c-luciferases evolving under episodic diversifying selection are also correlated significantly with the rate of decay of light, suggesting selection diversified enzyme kinetics. Of particular interest is site 160 because it is one of only a handful of sites that differs between *Photeros annecohenae* and *P. morini*. Despite strong similarity in c-luciferases in these two species, their enzymes have dramatic differences in measures of decay rate (Fig. 1) that are related to duration of courtship pulses (20). However, which selective forces influence rates of light decay at the organismal level is not yet certain because there are very few experiments on the fitness effects of different ostracod signals (47). We hypothesize pulse duration, in part dictated by enzyme kinetics (20), may be important for fitness via inter-specific anti-predator displays and/or through mate recognition or choice. Consistent with this hypothesis, there is extensive variation among species, with pulse durations in courtship signals ranging from ~0.15 - 9 seconds across species (20, 24). In addition to sexual selection, c-luciferase kinetics could also be under selection at the organismal level due to changes in environmental factors like temperature, pH, or salinity. Because cypridinid luciferase is secreted externally, and bioluminescent ostracods are globally distributed, environmental factors affecting enzyme function vary widely. Indeed, previous *in vitro* expression of *C. noctiluca* luciferase showed temperature-dependent differences in activity (32, 48). Increased sampling of taxa living in different temperatures, including deep-sea cypridinids, could allow testing of varying conditions and the role that adaptation may play in constraining or facilitating changes in rates of light decay.

Another possibility is that variation in some bioluminescent functions are selectively neutral, which seems to be true for color, including emission width (FWHM) and perhaps λ_{max}. We see no clear pattern in variation of FWHM, and no correlation of this parameter with any positively selected sites in c-luciferases. Although we do not yet have good candidate mutations for the genetic basis of FWHM, we do have such candidates for λ_{max}, thanks to mutagenesis experiments (45). One site (178) strongly affects λ_{max} and has a dn:ds ratio indistinguishable from one, consistent with a minimal impact on fitness, suggesting it evolved neutrally. In

contrast, three sites (74, 114, 160) are correlated with changes in λ_{max} and under positive selection, which could indicate selection drove some changes in color. The patterns of amino acid differences in these selected sites also show differences between non-courtship and courtship-signaling species. At the organismal level, λ_{max} is non-random with respect to phylogeny (Fig. 1, 3) and we observe evolutionary shifts in λ_{max} between non-courtship and courtship-signaling species and separately, along the branch leading to *Photeros*. If the color change between non-courtship and courtship-signaling species is robust to greater taxon sampling, it could be adaptive, perhaps linked to environmental differences or differences in predators' vision. In contrast, the color change in *Photeros* does not seem adaptive. Because *Photeros* are among the few signaling species that live in seagrass (22), the "reflectance hypothesis" (49) could predict the ancestor of *Photeros* underwent an adaptive shift to a greener signal to increase signal efficacy, with a correlated shift to grass bed habitat. However, even *Photeros* species that live in non-grass habitats (like *P. morini*) have green-shifted λ_{max} (Fig. 1, 3). Furthermore, an adaptive green-shift is not supported by patterns of substitution in c-luciferases because none of the positively selected sites are fixed in *Photeros*, while also different from non-*Photeros*. Finally, we did not find evidence of positive selection specifically along the branch leading to *Photeros.* If λ_{max} and FWHM are in fact neutral for organismal fitness, other factors like constraints could instead dictate changes we observed in color, especially λ_{max}.

Biochemical constraints like pleiotropy are likely to influence the evolution of bioluminescent phenotypes. One site in c-luciferase (160) is under positive selection and correlated to both λ_{max} and light decay. This suggests a possible role of pleiotropy in phenotypic evolution if mutations in c-luciferase affect multiple phenotypic parameters at once. However, counter to a pervasive role for pleiotropy, we do not find a strong relationship between λ_{max} and light decay constant across all species in our study (Fig. S3). Instead, all *Photeros* have similar λ_{max} while rates of light decay vary considerably between those species. Still, we cannot fully rule out biochemical constraint as a driver for the evolution of emission spectra because such constraints may have changed during evolution, for example during the shift in λ_{max} in early *Photeros*. If so, modern *Photeros* could have biochemical constraints that differ from ancestral species, now allowing rates of light decay to change independently of λ_{max}. Testing an evolutionary change in biochemical constraint would entail extensive mutagenesis and expression experiments guided by reconstructing the history of *Photeros* luciferase. Constraints may also arise from epistatic interactions among sites, of which we find some evidence in structuring phenotypic differences between species despite lacking statistical power to exhaustively cover all site-by-site interactions. As phenotypes like color and/or decay evolve, previous changes at certain sites (such as 404) will influence the magnitude of new mutations on function. It is also possible that site-specific epistatic interactions changed during the evolution of c-luciferases, like in other proteins (50, 51).

Taken together, our results support the hypothesis that molecular evolution of c-luciferase impacted the diversity of signal phenotypes in the sea firefly radiation of species. The results further illustrate the potential for varied interactions between molecular evolution, pleiotropy, epistasis, and phenotypes during diversification - even when considering a single gene like c-luciferase. When extrapolated to other genes, such as the presumably many genes

affecting behavioral phenotypes and in different environments, the combinatorics become astronomical, providing new perspectives on how life became so incredibly diverse. Such a pluralistic view of evolution, incorporating many different processes at different levels of organization (52, 53), allows for a more holistic understanding of how biodiversity originates.

Materials and Methods

All data and code for all analyses are available on GitHub: https://github.com/ostratodd/Cypridinidae_Emission.

Specimen Collection. We collected animals using small nets while SCUBA diving, or by setting baited traps, as previously described (54). We collected multiple species from each of four Caribbean locations: Discovery Bay, Jamaica; Bocas del Toro, Panama; South Water Caye, Belize; and Roatan, Honduras. We also analyzed one species from Catalina Island, California and two from Isla Magueyes, Puerto Rico. We purchased dried *Vargula hilgendorfii*, which originates in Japan (Carolina Biological). We report specific collection localities and size measurements (SI, Table S5). For transcriptomes, we preserved specimens in RNALater (Invitrogen). We carried some animals alive to UCSB to measure emission spectra and others we dried using different methods depending on what tools were available on-site. To induce bioluminescence, we crushed live or dried specimens in seawater using a small plastic pestle, usually one animal at a time.

Luciferase discovery and amplification. We discovered 12 putative luciferase genes from transcriptomes of cypridinid ostracods. To find these, we queried a previously published transcriptome of *V. tsujii* (55), and new transcriptomes using whole bodies of luminous cypridinids. We will use these transcriptomes for a future phylogenetic study, so we present here paired-end transcriptomes containing putative luciferases, submitted to BioProject PRJNA589015. Metadata from sequencing and NCBI accession number are included in (Table S5).

For six species, we amplified putative c-luciferase sequences via PCR and confirmed nucleotide sequences with Sanger sequencing. In many cases, we found complete assembly of putative luciferase genes from transcriptomes depends on which assembly programs and parameters we used. We believe the VWD domains contained in luciferases (38) have duplicated wildly in ostracods, leading to challenges when assembling transcriptomes from multiple individuals. In some cases, the final assembly in our BioProject differs from early assemblies we used to design primers in calling different regions of a single luciferase gene different isoforms. We analyzed the Sanger-sequenced genes and provide a multiple sequence alignment of corresponding sequences (Fig. S4).

Luciferase Expression In Vitro. We expressed putative c-luciferases from exemplars of four major clades of cypridinids, *Maristella*, *Photeros*, *Kornickeria*, and *Vargula*. Using mammalian cells, we expressed three putative c-luciferase proteins (*Maristella sp. SVU*, *Vargula tsujii* and *Photeros morini*) and in *Pichia* yeast, we expressed three putative luciferases (*V. tsujii* and *M.* sp. “SVU”, and *Kornickeria hastingsi*). We also expressed previously published *Cypridina noctiluca* luciferase (32) in yeast as a positive control. These proteins are secreted

into culture media, to which we added luciferin substrate to test the hypothesis that they are luciferases and catalyze a light reaction. For more details, se SI.

Light catalysis assay. To assess the ability of constructs to catalyze a light reaction, we harvested cell culture media from mammalian or yeast cells from each transfection (approximately 10 - 25 mL per transfection), and concentrated it using 30,000 MWCO centrifugal filters (Amicon), spun from 30-240 minutes at 4,000 x g. After centrifugation, the protein solution was immediately collected for the assay. Varying volumes of concentrated protein solution and luciferin assay mix (Targeting Systems; prepared to manufacturer's specifications but with unknown concentration for mammalian cells; and for yeast, we procured vargulin from NanoLight Technology (Pinetop, AZ) and suspended it in 10 mM TRIS to a working concentration of 0.01 ng/ul) were added together in a plate reader (Wallac). We measured luminescence in counts per second (CPS) for 10 seconds.

Phylogeny of luciferases. We generated a gene tree of translated luciferase candidates and closely related genes. We used a published c-luciferase from *V. hilgendorfii* as a query in a similarity search using BLAST, retaining the top 20 hits from each transcriptome searched. In some cases, we did not obtain full-length luciferase transcripts in the assembly created by Trinity (56), which we attribute to polymorphism from pooling individuals. In these cases, we did a second assembly of luciferase fragments using cap3 (57) and selected the longest orfs as putative c-luciferases. We aligned these sequences using MAFFT (58). We used IQTREE (59) to select the best-fit model of protein evolution and to estimate the maximum likelihood phylogeny. We rooted this phylogeny using midpoint rooting to identify putative c-luciferases from new transcriptomes as orthologues to previously published c-luciferases. Both *P. annecohenae* and *P. sp. WLU* had two additional genes (in-paralogs) in this clade, whereas all other species had one direct ortholog of published c-luciferases in the clade. We excluded these two in-paralogs from further analysis because we did not confirm these transcriptome sequences with PCR and because the in-paralog sequence from *P. annecohenae* lacks a signal peptide and is therefore unlikely to be a functional c-luciferase.

Testing for positive selection in luciferases. We aligned luciferase DNA by codon, first aligning amino acids in MAFFT (58), then matching DNA codons using pal2nal (60). We used Hyphy (61) to compare ratios of synonymous to non-synonymous substitutions. We used MEME and FEL to search for individual codons under diversifying selection (39).

Emission Spectra. We used a home-built fluorimeter (spectroradiometer) at UCSB to measure emission spectra. We imaged the output orifice of the integrating sphere or a cuvette cell's transparent wall by a relay lens onto the entrance slit of a spectroradiometer (Acton SpectraPro 300i) equipped with a charge-coupled device camera (CCD) detector (Andor iDus). Because of the limited time-span of bioluminescence, we collected a series of data frames upon introduction of specimens. We usually sampled at 10 time points every two seconds for each emission with 2 seconds integration time, although sample numbers ranged from 5-20, depending on the species. The spectral data acquired by the CCD camera were corrected for instrumental response artifacts by measuring the spectrum of a black body-like light source (Ocean Optics LS-1) and calculating appropriate correction factors. We also took background spectra for each experiment to subtract from the experimental emission spectra (see SI for further details). For the high quality spectra, we employed Savitzky-Golay smoothing using the

signal package (62) in R (63) and then calculated λ_{max} (the wavelength with the highest emission value) and Full Width Half Max (FWHM, the width of the spectrum in nanometers where emission is half the value of maximum).

Genetic Basis of Changes in c-luciferase Function. We looked for amino acid changes associated with changes in three functions: λ_{max}, FWHM, and light decay constant. To find mutations in luciferase that shift λ_{max}, we analyzed previously published data from mutagenesis experiments on *Cypridina noctiluca* luciferase (*Cn-luc*) and data on luciferases from 9 other cypridinid species with both luciferase and emission spectrum data. Kawasaki et al. (45) created 35 variants of *Cn-luc* and measured λ_{max} for each variant plus wild type *Cn-luc*. To measure λ_{max} they added luciferin to *Cn-luc* variants expressed heterologously, finding λ_{max} to vary between 435 nm and 463 nm. They did not report complete spectra or FWHM for most of the mutants, instead only reporting λ_{max}. Therefore, our subsequent analyses of FWHM use only data from natural c-luciferase sequences and our newly reported FWHM data from emission spectra. For the mutational analysis, each variant contained 1-8 mutated sites compared to wild type, and across all 35 variants, a total of 23 different amino acid sites contained mutations. For the non-*Cn-luc* luciferases, we aligned sequences to *Cn-Luc* using MUSCLE (64) to identify amino acids at sites homologous to those mutated in *Cn-Luc* variants. We refer to sites as numbered for the homologous positions in *Cn-luc*, which differs from their aligned position. For light decay constant, we used previously published data (20).

We analyzed candidate mutations with ANOVA to test for significant associations between variant amino acid sites and functions of c-luciferases, similar to methods of Yokoyama et al. for opsins and absorbances (46). To determine which sites were most highly correlated with each function, we used a model selection approach in the R package MuMIn (ref). We used the dredge function to test combinations of amino acid sites, which formed different models. Dredge sorts models by AIC score. We tallied amino acid sites present in the highest number of best-fit models, and then performed ANOVA using a model with those sites.

Acknowledgements

We thank C. Pham, T. Halvorsen, Y. Chiu, J. Russo, M. Pfau, and T. Jessup for assistance with cloning and expression. We acknowledge J. Lachat and Y. Nishimiya for assistance with the *Pichia* system and for providing the *C. noctiluca* construct used as a positive control. We are grateful to Z. Terner from the UCSB Statistical Consulting Laboratory for his advice on modeling. We thank K. Niwa for assistance in the early stages of emission spectra collection. We also extend our thanks to Y J. Li, S. Schulz, N. Reda, J. Morin, M D. Ramirez, C. Motta, and V. Gonzalez for assistance with animal collections. Organisms were collected under the auspices of proper permitting from the Jamaican National Environment and Planning Agency (Permit Ref. #18/27), the Belize Fisheries Department (Permit #000003-16), the Honduran Department of Fish and Wildlife (Permit #DE-MO-082-2016), the Puerto Rican Department of Natural and Environmental Resources (DRNA; Permit #2016-IC-113), and Panamanian Ministry of the Environment (MiAMBIENTE; Permit #SE/A-33-17). NMH was funded in part by an NSF EAPSI Fellowship (#1713975), the NSF Graduate Research Fellowship Program, and the UCSB Research Mentorship Program run by Dr. Lina Kim. NSF DDIG #1702011 to EAE provided funding for transcriptomes and emission spectra estimation. JC received funding from the UCSB URCA program. ET was funded by a CSUPERB Research

Development Grant for transcriptomes, gene synthesis, and vector cloning. Authors THO, GAG, and ET were funded by NSF (#DEB-1457754, DEB-1457439, and DEB-1457462 respectively.)

References Cited

1. T. F. Hansen, The Evolution of Genetic Architecture. *Annu. Rev. Ecol. Evol. Syst.* **37**, 123–157 (2006).

2. T. F. Mackay, The genetic architecture of quantitative traits. *Annu. Rev. Genet.* **35**, 303–339 (2001).

3. J. Flint, T. F. C. Mackay, Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res.* **19**, 723–733 (2009).

4. A. B. Lall, H. H. Seliger, W. H. Biggley, J. E. Lloyd, Ecology of Colors of Firefly Bioluminescence. *Science* **210**, 560–562 (1980).

5. E. A. Ellis, T. H. Oakley, High Rates of Species Accumulation in Animals with Bioluminescent Courtship Displays. *Curr. Biol.* **26**, 1916–1921 (2016).

6. P. Brand, *et al.*, The evolution of sexual signaling is linked to odorant receptor tuning in perfume-collecting orchid bees. *Nat. Commun.* **11**, 244 (2020).

7. R. Nagel, F. Kirschbaum, R. Tiedemann, Electric organ discharge diversification in mormyrid weakly electric fish is associated with differential expression of voltage-gated ion channel genes. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **203**, 183–195 (2017).

8. T. C. Mendelson, K. L. Shaw, Sexual behaviour: rapid speciation in an arthropod. *Nature* **433**, 375–376 (2005).

9. D. L. Stern, V. Orgogozo, The loci of evolution: how predictable is genetic evolution? *Evolution* **62**, 2155–2177 (2008).

10. C. K. Ellison, C. Wiley, K. L. Shaw, The genetics of speciation: genes of small effect underlie sexual isolation in the Hawaiian cricket Laupala. *J. Evol. Biol.* **24**, 1110–1119 (2011).

11. A. R. Templeton, Mechanisms of Speciation - A Population Genetic Approach. *Annu. Rev. Ecol. Syst.* **12**, 23–48 (1981).

12. S. Wright, *Evolution and the Genetics of Populations, Volume 3: Experimental Results and Evolutionary Deductions* (University of Chicago Press, 1984).

13. M. Supple, R. Papa, B. Counterman, W. Owen McMillan, The Genomics of an Adaptive Radiation: Insights Across the Heliconius Speciation Continuum. *Advances in Experimental Medicine and Biology*, 249–271 (2014).

14. D. Brawand, *et al.*, The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**, 375–381 (2014).

15. J. Spaethe, A. D. Briscoe, Early duplication and functional diversification of the opsin gene family in insects. *Mol. Biol. Evol.* **21**, 1583–1594 (2004).

16. M. J. Henze, T. H. Oakley, The Dynamic Evolutionary History of Pancrustacean Eyes and Opsins. *Integr. Comp. Biol.* **55**, 830–842 (2015).

17. S. Yokoyama, Molecular genetic basis of adaptive selection: examples from color vision in vertebrates. *Annu. Rev. Genet.* **31**, 315–336 (1997).

18. S. Yokoyama, *et al.*, Adaptive evolutionary paths from UV reception to sensing violet light by epistatic interactions. *Sci Adv* **1**, e1500162 (2015).

19. T. C. Mendelson, M. D. Martin, S. M. Flaxman, Mutation-order divergence by sexual selection: diversification of sexual signals in similar environments as a first step in speciation. *Ecol. Lett.* **17**, 1053–1066 (2014).

20. N. M. Hensley, *et al.*, Phenotypic evolution shaped by current enzyme function in the bioluminescent courtship signals of sea fireflies. *Proceedings of the Royal Society B* **286**, 20182621 (2019).

21. G. A. Gerrish, J. G. Morin, Living in sympatry via differentiation in time, space and display characters of courtship behaviors of bioluminescent marine ostracods. *Mar. Biol.* **163**, 190 (2016).

22. J. G. Morin, Luminaries of the reef: The history of luminescent ostracods and their courtship displays in the Caribbean. *J. Crustacean Biol.* **39**, 227–243 (2019).

23. T. J. Rivers, J. G. Morin, Complex sexual courtship displays by luminescent male marine ostracods. *J. Exp. Biol.* **211**, 2252–2262 (2008).

24. A. C. Cohen, J. G. Morin, Two New Bioluminescent Ostracode Genera, Enewton And Photeros (Myodocopida: Cypridinidae), with Three New Species from Jamaica. *J. Crustacean Biol.* **30**, 1–55 (2010).

25. J. G. Morin, Firefleas of the Sea: Luminescent Signaling in Marine Ostracode Crustaceans. *Fla. Entomol.* **69**, 105–121 (1986).

26. G. A. Gerrish, J. G. Morin, T. J. Rivers, Z. Patrawala, Darkness as an ecological resource: the role of light in partitioning the nocturnal niche. *Oecologia* **160**, 525–536 (2009).

27. E. N. Harvey, Studies on bioluminescence. XVI. What determines the color of the light of luminous animals. *Am. J. Physiol.* (1924).

28. K. P. Stevens, STUDIES ON THE AMOUNT OF LIGHT EMITTED BY MIXTURES OF CYPRIDINA LUCIFERIN AND LUCIFERASE. *J. Gen. Physiol.* **10**, 859–873 (1927).

29. A. L. Huvard, Analysis of visual pigment absorbance and luminescence emission spectra in marine ostracodes (Crustacea: Ostracoda). *Comp. Biochem. Physiol. A Physiol.* **104**, 333–338 (1993).

30. F. I. Tsuji, R. V. Lynch, C. L. Stevens, Properties of luciferase from the bioluminescent

crustacean, Cypridina hilgendorfii. *Biochemistry* **13**, 5204–5209 (1974).

31. E. A. Widder, M. I. Latz, J. F. Case, MARINE BIOLUMINESCENCE SPECTRA MEASURED WITH AN OPTICAL MULTICHANNEL DETECTION SYSTEM. *Biol. Bull.* **165**, 791–810 (1983).

32. Y. Nakajima, K. Kobayashi, K. Yamagishi, T. Enomoto, Y. Ohmiya, cDNA cloning and characterization of a secreted luciferase from the luminous Japanese ostracod, Cypridina noctiluca. *Biosci. Biotechnol. Biochem.* **68**, 565–570 (2004).

33. Y. Tochigi, *et al.*, Sensitive and convenient yeast reporter assay for high-throughput analysis by using a secretory luciferase from Cypridina noctiluca. *Anal. Chem.* **82**, 5768–5776 (2010).

34. A. M. Dean, J. W. Thornton, Mechanistic approaches to the study of evolution: the functional synthesis. *Nat. Rev. Genet.* **8**, 675–688 (2007).

35. T. R. Fallon, *et al.*, Firefly genomes illuminate parallel origins of bioluminescence in beetles. *Elife* **7** (2018).

36. V. R. Viviani, D. Amaral, R. Prado, F. G. C. Arnoldi, A new blue-shifted luciferase from the Brazilian Amydetes fanestratus (Coleoptera: Lampyridae) firefly: molecular evolution and structural/functional properties. *Photochem. Photobiol. Sci.* **10**, 1879–1886 (2011).

37. E. M. Thompson, S. Nagata, F. I. Tsuji, Cloning and expression of cDNA for the luciferase from the marine ostracod Vargula hilgendorfii. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 6567–6571 (1989).

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

39. B. Murrell, *et al.*, Detecting Individual Sites Subject to Episodic Diversifying Selection. *PLoS Genetics* **8**, e1002764 (2012).

40. S. L. Kosakovsky Pond, S. D. W. Frost, Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* **22**, 1208–1222 (2005).

41. V. R. Viviani, Y. Ohmiya, Bovine serum albumin displays luciferase-like activity in presence of luciferyl adenylate: insights on the origin of protoluciferase activity and bioluminescence colours. *Luminescence* **21**, 262–267 (2006).

42. A. C. Cohen, J. G. Morin, Sexual Morphology, Reproduction and the Evolution of Bioluminescence in Ostracoda. *The Paleontological Society Papers* **9**, 37–70 (2003).

43. J. G. Morin, A. C. Cohen, A guide to the morphology of bioluminescent signaling cypridinid ostracods from the Caribbean Sea, and a tabular key to the genera. *Zootaxa* **4303**, 301–349 (2017).

44. N. J. Reda, *et al.*, Maristella, a new bioluminescent ostracod genus in the Myodocopida (Cypridinidae). *Zool. J. Linn. Soc.* (2019) https:/doi.org/10.1093/zoolinnean/zlz082

(November 8, 2019).

45. K. Kawasaki, Y. Morita, S. Ohgiya, Y. Ohmiya, Y. Ohyama, Mutant luciferase. *US Patent* (2012) (August 29, 2016).

46. S. Yokoyama, H. Yang, W. T. Starmer, Molecular basis of spectral tuning in the red- and green-sensitive (M/LWS) pigments in vertebrates. *Genetics* **179**, 2037–2043 (2008).

47. T. J. Rivers, J. G. Morin, Female ostracods respond to and intercept artificial conspecific male luminescent courtship displays. *Behav. Ecol.* **24**, 877–887 (2013).

48. R. Yasuno, Y. Mitani, Y. Ohmiya, Effects of N-Glycosylation Deletions on Cypridina Luciferase Activity. *Photochem. Photobiol.* **94**, 338–342 (2018).

49. J. A. Endler, Signals, Signal Conditions, and the Direction of Evolution. *Am. Nat.* **139**, S125–S153 (1992).

50. E. A. Ortlund, J. T. Bridgham, M. R. Redinbo, J. W. Thornton, Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* **317**, 1544–1548 (2007).

51. S. Yokoyama, *et al.*, Epistatic adaptive evolution of human color vision. *PLoS Genet.* **10**, e1004884 (2014).

52. A. Seilacher, ARBEITSKONZEPT ZUR KONSTRUKTIONS-MORPHOLOGIE. *Lethaia* **3**, 393–396 (1970).

53. N. Tinbergen, On aims and methods of ethology. *Zeitschrift für tierpsychologie* **20**, 410–433 (1963).

54. A. C. Cohen, T. H. Oakley, Collecting and processing marine ostracods. *J. Crustacean Biol.* (2017).

55. T. H. Oakley, J. M. Wolfe, A. R. Lindgren, A. K. Zaharoff, Phylotranscriptomics to Bring the Understudied into the Fold: Monophyletic Ostracoda, Fossil Placement and Pancrustacean Phylogeny. *Mol. Biol. Evol.* (2012) https:/doi.org/10.1093/molbev/mss216.

56. B. J. Haas, *et al.*, De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).

57. X. Huang, A. Madan, CAP3: A DNA sequence assembly program. *Genome Res.* **9**, 868–877 (1999).

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

59. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

60. M. Suyama, D. Torrents, P. Bork, PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–12

(2006).

61. S. L. K. Pond, *et al.*, HyPhy 2.5—A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. *Molecular Biology and Evolution* (2019) https:/doi.org/10.1093/molbev/msz197.

62. U. Ligges, *et al.*, signal: Signal processing. *Retrieved from Available at: https://CRAN. R-project. org/package= signal* (2015).

63. R. C. Team, Others, R: A language and environment for statistical computing (2013).

64. R. C. Edgar, MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).

Ellis, EA. 2019. "Macroevolutionary Consequences of Rapid Species Accumulation due to Sexual Selection." Order No. 13901546, University of California, Santa Barbara.

Torres E**,** Gonzalez VL. 2007. Molecular phylogeny of cypridinid ostracodes and the evolution of bioluminescence. In: Proceedings of the 14th International Symposium on Bioluminescence and Chemiluminescence: Chemistry, Biology, and Applications. Szalay A.A., P.J. Hill, L.J. Kricka, P.E. Stanley, eds. New Jersey: World Scientific Publishing Company, p. 269-272.

Link to Tables (In Spreadsheet Format)

Link to supplemental figures and legends

Author information (contributions as here):

1. Nicholai Marcus Hensley
    - Affiliation: Department of Ecology, Evolution, & Marine Biology, University of California, Santa Barbara
    - Contribution: Conceptualization, Funding Acquisition, Methodology, Investigation, Writing - Review & Editing, Project Administration, Data Curation, Formal Analysis, Software
    - Competing interests: No competing interests declared
    - ORCID: 0000-0003-0681-0808
2. Emily A. Ellis
    - Affiliation: Florida Museum of Natural History, University of Florida
    - Contribution: Conceptualization, Funding Acquisition, Methodology, Investigation, Writing - Review & Editing
    - Competing interests: No competing interests declared
3. Nicole Y. Leung
    - Affiliation: Neuroscience Research Institute and Department of Molecular, Cellular and Developmental Biology, University of California, Santa Barbara California
    - Contribution: Investigation, Writing - Review & Editing

- Competing interests: No competing interests declared
- ORCID: 0000-0001-8823-1783

4. John Coupart
    - Affiliation: Department of Ecology, Evolution, & Marine Biology, University of California, Santa Barbara
    - Contribution: Investigation
    - Competing interests: No competing interests declared
5. Alexander Mikhailovsky
    - Affiliation: Department of Chemistry and Biochemistry, University of California, Santa Barbara
    - Contribution: Methodology, Resource, Writing - Review & Editing
    - Competing interests: No competing interests declared
6. Daryl A. Taketa
    - Affiliation: Department of Molecular, Cellular and Developmental Biology, University of California, Santa Barbara
    - Contribution: Methodology, Investigation
    - Competing interests: No competing interests declared
7. Michael Tessler
    - Affiliation: American Museum of Natural History and New York University, New York
    - Contribution: Resources, Writing - Review & Editing
    - Competing interests: No competing interests declared
    - ORCID: 0000-0001-7870-433X
8. David F. Gruber
    - Affiliation: Department of Biology and Environmental Science, City University of New York Baruch College, New York
    - Contribution: Funding Acquisition, Resources, Writing - Review & Editing
    - Competing interests: No competing interests declared
    - ORCID: 0000-0001-9041-2911
9. Anthony W. De Tomaso
    - Affiliation: Department of Molecular, Cellular and Developmental Biology, University of California, Santa Barbara
    - Contribution: Supervision, Resources
    - Competing interests: No competing interests declared
10. Yasuo Mitani
    - Affiliation: Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Sapporo
    - Contribution: Funding Acquisition, Supervision, Writing - Review & Editing, Resources
    - Competing interests: No competing interests declared
11. Trevor J. Rivers
    - Affiliation: Department of Ecology and Evolutionary Biology, University of Kansas
    - Contribution: Resources

- Competing interests: No competing interests declared

12. Gretchen A. Gerrish
    - Affiliation: Trout Lake Station, Center for Limnology, University of Wisconsin - Madison
    - Contribution: Funding Acquisition, Project Administration, Resources, Writing - Review & Editing
    - Competing interests: No competing interests declared
    - ORCID: 0000-0001-6192-5734
13. Elizabeth Torres
    - Affiliation: Department of Biological Sciences, California State University Los Angeles
    - Contribution: Funding Acquisition, Resources, Project Administration, Writing - Review & Editing
    - Competing interests: No competing interests declared
14. Todd Hampton Oakley
    - Affiliation: Department of Ecology, Evolution, & Marine Biology, University of California, Santa Barbara
    - Contribution: Conceptualization, Visualization, Writing – Original Draft Preparation, Writing - Review & Editing, Software, Data Curation, Funding Acquisition, Formal Analysis, Supervision, Project Administration, Resources
    - Competing interests: No competing interests declared
    - ORCID: 0000-0002-4478-915X

Supplemental Materials and Methods

Specimen collection

We identified most species by the patterns of male courtship signals, targeting each particular display type with a single net and later measuring length and height of individual animals using a dissecting microscope and ocular micrometer. Several of the species, especially from Panama, are undescribed and we refer to those here by field codes, consisting of two or three capital letters. We classify species into genera based on length:height ratio, which is a reliable genus-level characteristic (1, 2). For emission spectra specimen preservation, methods varied by locale. In Roatan and Puerto Rico, we dried ostracods in direct sunlight. In Panama and Belize, we used a drying oven and transported animals in Eppendorf tubes with silica beads as a dessicant.

Luciferase discovery and amplification.

We designed luciferase-specific primers (Table S6) to amplify from cDNA and obtain sequences that do not include signal peptides (18 or 19 amino acids from the n-terminus, as inferred with SignalP) because we later used yeast-specific signal peptides during protein expression. we used an initial denaturation of 95°C for 2 min. For 30x cycles, we performed a 95°C denaturation step for 1 min., followed by an annealing phase at varying temperatures per species (*K. hastingsi*: 45.5°C, *P. morini*: 48°C, *M. sp.* SVU: 41.1°C , *M. sp.* SVD: 43.7°C, *M chicoi*: 41.4°C, *V. tsujii*: 45.5°C) for 1:45 min., and then by an extension step at 73°C for 1 min. For *V. tsujii* primers designed from the published transcriptome, we used thermal profile: 40 cycles of 94°C for 35s, 55°C for 30s, 72°C for 1min) and amplified the native signal peptide.

Luciferase Expression In Vitro.

We expressed three luciferases in mammalian HEK293 cells. To construct a *V. tsujii* luciferase (VtL) expression vector, we first amplified VtL using primers with engineered restriction sites to clone into a pCR4-TOPO vector. We next excised VtL-pCR4 with XhoI and EcoRI (Promega), and subcloned into a modified pCMV3B mammalian expression vector with a C-terminal mCherry reporter (mCherry-C). The luciferase genes of *P. morini* and *M. sp. SVU* were synthesized and cloned into the mCherry-C vector by Genscript (Piscataway, New Jersey, USA) with flanking restriction enzyme sites. We planned to use mCherry to quantify the concentration of expressed luciferase, but we found high autofluorescence of cell media and/or other secreted proteins to preclude this use. We first transformed cloned constructs into competent *E. coli* cells using the One Shot Chemical Transformation Kit (Invitrogen), and cultured for 24 hours in standard lysogeny broth (LB) with 0.1% kanamycin at 37°C. We verified construct transformation using the engineered restriction enzyme sites in digests and compared them to their expected product size. We extracted these plasmids using the FastPlasmid Mini kit (Qaigen) and assessed concentrations with the Qubit high standard DNA kit (Qubit). For transfection, we cultured mammalian HEK293 cells in Dulbecco's modified Eagle's medium (DMEM), supplemented with 10% fetal bovine serum (FBS) and penicillin/streptomycin (P/S) at 37°C with 5% CO_2. We then plated 5 x 10^4 cells in each well of a 24-well plate one to three days before transfection. Cell medium was changed to DMEM without serum and antibiotics before

transfection. We transfected cells with 0.5 µg of vector using Lipofectamine 2000 (Invitrogen), performed according to the manufacturer's instructions. After 4 hours of incubation, we replaced the transfection medium with DMEM+FBS+P/S and allowed the cells to recover for 24 hours. We collected cells via trypsin digestion and reseeded them into 10mL of DMEM+FBS+P/S+1% G418 to select against untransfected cells in 90cm cell plates. We cultured the transfected cells for 3 to 5 days before harvesting and using in light catalysis assays.

For expression in *Pichia* yeast, we cloned sequences into the pPICZ-αC vector at the XhoI and NotI sites following standard procedures (Invitrogen Easy Select *Pichia* kit). First, we analyzed predicted c-luciferases for the presence of a signal peptide at the n-terminal end using SignalP v4.1 (3). We then designed primers for cloning and protein expression to amplify the entire c-luciferase sequence without the native signal peptide, beginning usually 51-54 bp inside the 5' end from the predicted start codon. 3' end primers excluded the native stop codon so that a fusion construct could be generated. Fusion constructs were made via the EasySelection Pichia expression kit (Invitrogen) using the pPICZ-αC vector according to the manufacturer's instructions. Briefly, we used the 5' XhoI site in order to generate fusion c-luciferases with an alpha secretion signal from yeast. We reconstructed the Kecx2 cleavage site with one Glycine Alanine repeat region via PCR. On the 3' end, we used NotI; this would result in the addition of extra amino acids in our expressed proteins on the c-terminal end before inclusion of the fusion c-myc epitope and histidine tags. We transformed newly-made, linearized constructs into *Pichia* using electroporation with a BioRad Micro-Pulser using the Sc2 program (1.5 V). After electroporation, *Pichia* were allowed to recover in selective media for an hour before plating. We initially selected for recombinant *Pichia* colonies using two concentrations of zeocin (100 and 500 mg/mL). After three days of growth, individual colonies were replica-plated at high zeocin concentrations (1,000 and 2,000 mg/mL) to try and screen for high copy-number integrants for our gene of interest. After one day, we selected single colonies that grew best at high zeocin concentrations to induce protein expression according to the manufacturer's guidelines. To stabilize the pH of the media for extended expression, colonies were grown in 25mL buffered media with glycerol in baffled flasks until the OD_{600} reached 2.0 - 8.0. For our colonies, this occurred after 72 hrs due to suboptimal shaking conditions. We then calculated the amount of original growth we would need for an OD_{600} of 1.0 in 30mL expression media, spun down the appropriate volume of the original colonies at 3,000 g for 5min., removed the glycerol media, and resuspended the pellet in 30mL of buffered media with methanol in a 125 mL baffled flask. Flasks were shaken in a table-top incubator at 29.5C at 300 rpm for 3 days, with media supplemented with 100% methanol every 24hrs to maintain a 0.5% volume of methanol in culture.

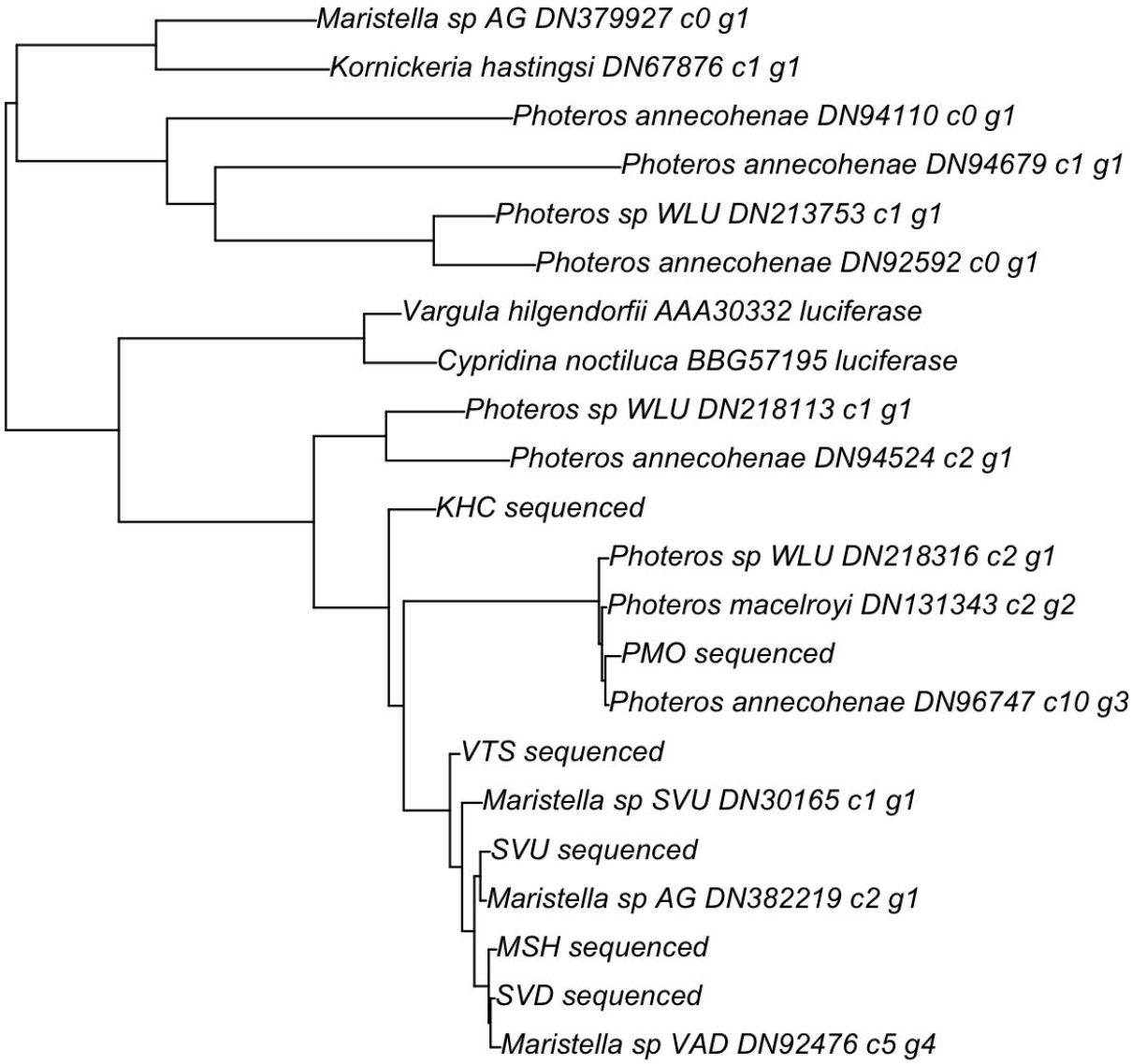
Emission Spectra

In earlier trials of emission spectra data collection, we introduced specimens into a test tube placed inside Spectralon-coated 150 mm diameter integrating sphere (Labsphere). But because we report relative rather than absolute levels of light at different wavelengths, we abandoned the integrating sphere in later trials for a rectangular quartz cuvette to increase sensitivity of our analyses.

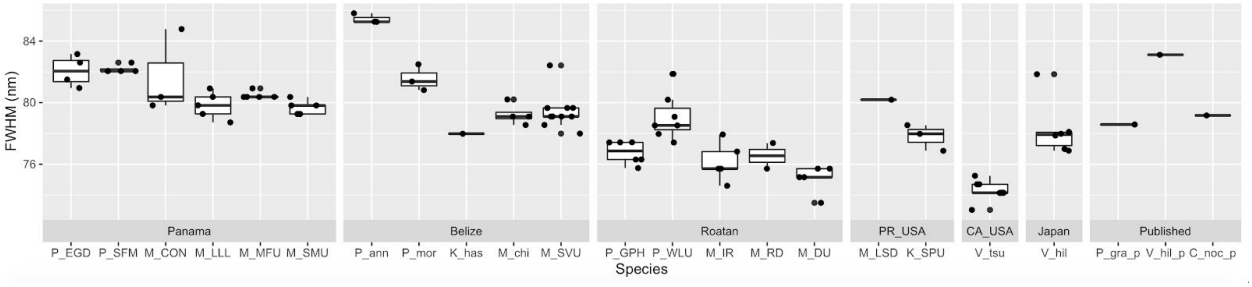
To correct for variation in emission, background noise, and quality during data collection, we first summed all collection time points for one sample, then subtracted each background value for each wavelength, and corrected using measurements from the black body radiator before standardizing each spectrum, setting the maximum value to 1.0 as the wavelength with the most photons. Some specimens did not yield strong light emission, probably due to variation in drying. We filtered low quality data using signal to noise ratio. Specifically, we sorted emission values at each wavelength from lowest to highest and averaged the lowest 1000 data points to estimate a baseline emission value (E_{min}). We then found the maximum emission value (E_{max}). From this, we calculated a signal to noise value as E_{min}/E_{max} , removing trials where $E_{min}/E_{max} < 0.02$.

Supplemental Figures and Tables

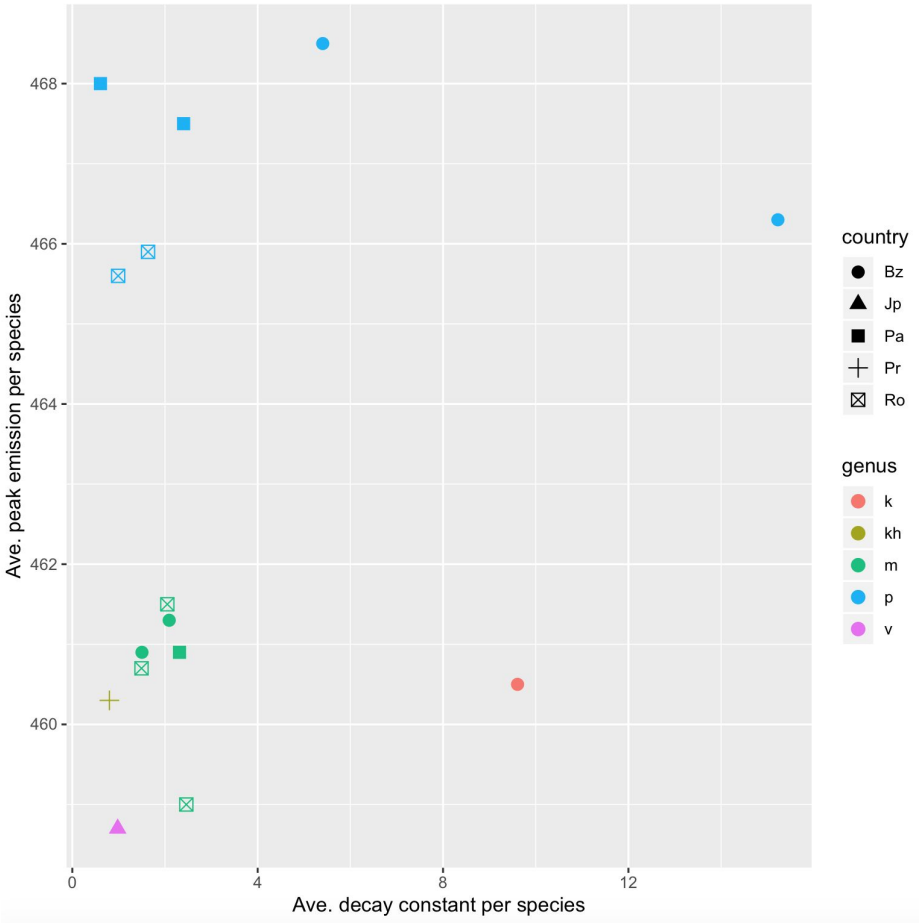
Supplemental Figure S1 - Maximum likelihood phylogeny of sequences most similar to published c-luciferases.



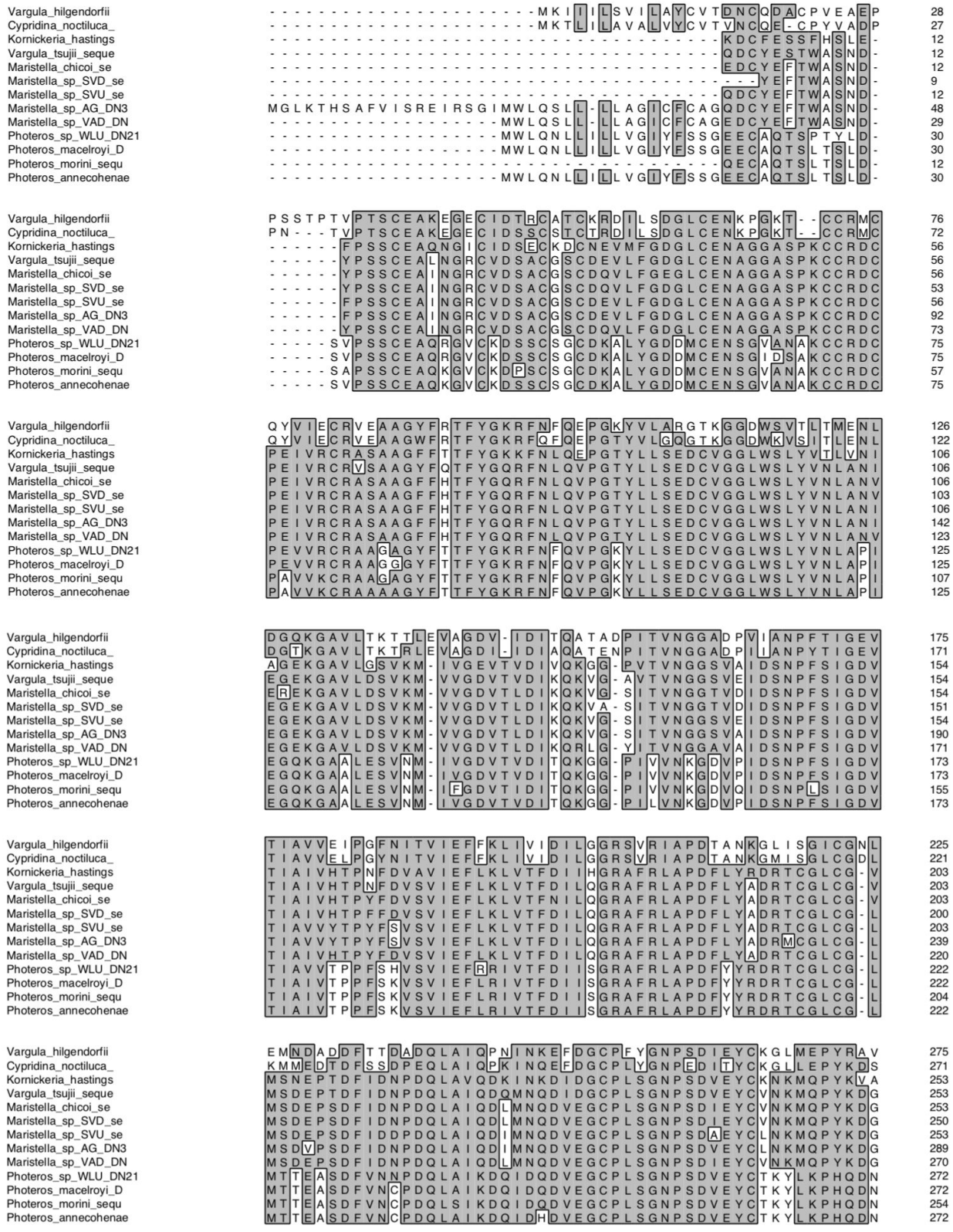
Supplemental Figure S2. Full Width of emission spectrum at Half of the Maximum value (FWHM) in nanometers (nm) from new emission spectra from 20 species and previously published spectra from 3 species.

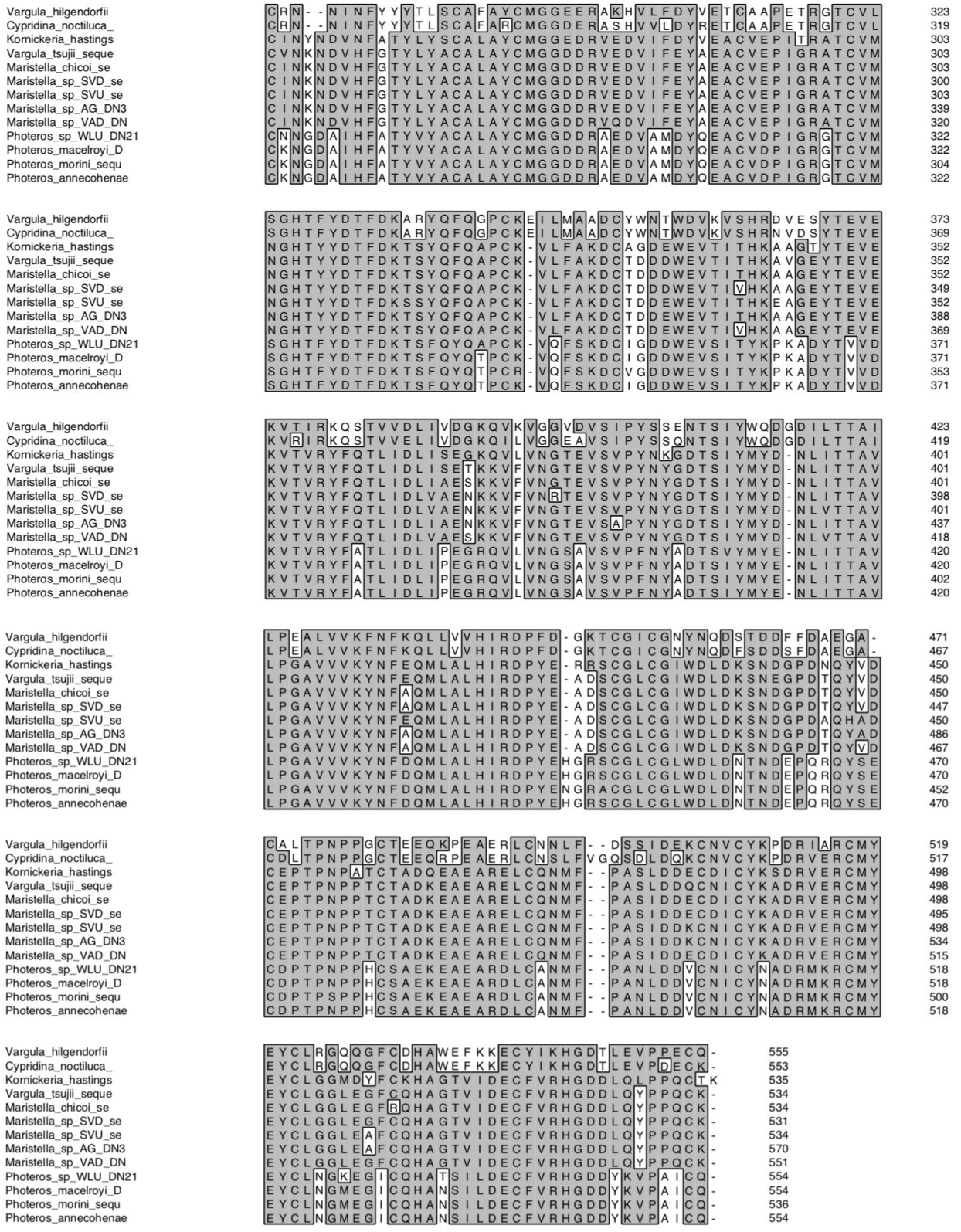


Supplemental Figure S3. We find no relationship between the color of light emission and light decay constants (Pearson's correlation test, t = , df = 16, p = 0.4). Scatter plot of the average decay (x-axis) and peak emission spectra (y-axis) measured for different species of luminescent ostracod (N = 17 species). Data are colored by genus, and shape by country of origin. Although *Photeros* greatly differ in their peak emission spectra than other genera, there is no strong pattern for differences in light decay constant.



Supplemental Figure S4. Multiple sequence alignment for putative and functional c-luciferase sequences





Supplemental 8

Table S1 - Previously published emission spectra for Cypridinidae

Species	λmax(nm)	FWHM (nm)	Method	Light Units	Citation
Vargula hilgendorfii	459		Absolute max	Relative Quanta	(4)
Vargula hilgendorfii	465	84	Savitzky-Golay with 2°polynomial and 25 channel smoothing, slit = 1, .1 mm	Energy	(5)
Vargula tsujii	466	87	Savitzky-Golay with 2°polynomial and 25 channel smoothing, slit=1 mm	Energy	(5)
Cypridina noctiluca	465		Not described	Not described	(6)
Cypridina noctiluca	454		Absolute Max following FFT/LPT (>0.05) smoothing in OriginPro. N=2	Not described	(7)
Photeros shulmanae	473	80	Cited (Widder et al. 1983)		(8)
Photeros graminicola	473	80	Cited (Widder et al. 1983)		(8)

Table S2 - Statistical results for luciferase expression *in vitro* from key comparisons

Culture	Comparison	Test	Value	Std. Error	DF	T-value	P-value	Correted P-value (if needed)
Mammalian Cells	HEK & P_mor	T-test	na	na	8.39	-4.76	0.00	0.00
Mammalian Cells	HEK & M_SVU	T-test	na	na	11.92	-14.91	0.00	0.00
Mammalian Cells	HEK & V_tsu	T-test	na	na	11.25	-15.92	0.00	0.00
Yeast Cells	Before & After substrate (C_noc)	T-test	na	na	8.75	5.79	0.00	0.00
Yeast Cells	Before & After substrate (K_has)	T-test	na	na	4.68	3.18	0.03	0.11
Yeast Cells	Before & After substrate (M_SVU)	T-test	na	na	11.50	2.62	0.02	0.09
Yeast Cells	Before & After substrate (V_tsu)	T-test	na	na	4.60	4.88	0.01	0.02
	Model effects							
Yeast Cells	Pichia after substrate addition	Linear Mixed Effect Model	2.13	0.85	53.00	2.50	0.02	
Yeast Cells	C_noc	Linear Mixed Effect Model	3.17	0.98	19.00	3.23	0.00	
Yeast Cells	K_has	Linear Mixed Effect Model	0.88	1.04	19.00	0.85	0.41	
Yeast Cells	M_SVU	Linear Mixed Effect Model	0.57	0.95	19.00	0.60	0.55	
Yeast Cells	V_tsu	Linear Mixed Effect Model	0.49	1.04	19.00	0.47	0.65	
Yeast Cells	Before substrate addition	Linear Mixed Effect Model	-0.09	0.28	40.00	-0.31	0.75	
Yeast Cells	C_noc * Before substrate addition	Linear Mixed Effect Model	-2.83	0.36	40.00	-7.87	0.00	
Yeast Cells	K_has * Before substrate addition	Linear Mixed Effect Model	-2.00	0.39	40.00	-5.09	0.00	
Yeast Cells	M_SVU * Before substrate addition	Linear Mixed Effect Model	-0.70	0.32	40.00	-2.19	0.03	
Yeast Cells	V_tsu * Before substrate addition	Linear Mixed Effect Model	-1.26	0.39	40.00	-3.21	0.00	

Table S3 - Means of parameters of emission spectra

Species abbreviation	N	Lmax_Mean	Lmax_SD	FWHM_Mean	FWHM_SD
P_EGD	4	468	0.71	82.05	1
P_SFM	4	467.5	0.71	82.19	0.28
M_CON	3	461.5	0.32	81.66	2.7
M_LLL	5	461	0.6	79.82	0.87
M_MFU	5	461.6	0.46	80.48	0.25
P_ann	3	468.5	1.8	85.44	0.32
P_mor	3	466.3	0.63	81.56	0.85
M_SMU	5	460.9	0.49	79.71	0.46
K_has	1	460.5	NA	77.98	NA
M_chi	4	461.3	0.55	79.24	0.7
M_SVU	10	460.9	0.85	79.43	1.2
P_GPH	6	465.6	0.78	76.77	0.74
P_WLU	7	465.9	1.5	79.08	1.5
M_LSD	1	459.4	NA	80.19	NA
M_IR	5	460.7	1.5	76.16	1.3
M_RD	2	461.5	2.7	76.54	1.2
M_DU	5	459	1.1	75.05	0.91
V_tsu	8	460.6	1.2	74.28	0.65
K_SPU	3	460.3	1.8	77.79	0.85
V_hil	6	458.7	2	78.28	1.8
P_gra_p	1	471.1	NA	78.58	NA
V_hil_p	1	460.5	NA	83.11	NA
C_noc_p	1	454.3	NA	79.17	NA

Table S4 - All parameters measured from individual emission spectra in this study, including those removed for further analysis due to low signal:noise

Abbre-viation	locality	genus	species	replicate	sex	preservation	source	sgMax	sgfwhm	error
P_EGD	Panama	Photeros	Photeros_EGD	EGD1	male	dried	ucsb	467.73	81.50	0.00
P_EGD	Panama	Photeros	Photeros_EGD	EGD3	male	dried	ucsb	467.18	82.60	0.00
P_EGD	Panama	Photeros	Photeros_EGD	EGD4	male	dried	ucsb	468.83	80.95	0.00
P_EGD	Panama	Photeros	Photeros_EGD	EGD5	male	dried	ucsb	468.28	83.16	0.00
P_SFM	Panama	Photeros	Photeros_SFM	SFM1	male	dried	ucsb	469.38	75.99	0.05
P_SFM	Panama	Photeros	Photeros_SFM	SFM2	male	dried	ucsb	467.18	82.05	0.00
P_SFM	Panama	Photeros	Photeros_SFM	SFM3	male	dried	ucsb	466.63	82.05	0.00
P_SFM	Panama	Photeros	Photeros_SFM	SFM4	male	dried	ucsb	467.73	82.60	0.00
P_SFM	Panama	Photeros	Photeros_SFM	SFM5	male	dried	ucsb	468.28	82.05	0.00
M_CON	Panama	Contragula	contragula	cont1	male	dried	ucsb	461.13	79.82	0.00
M_CON	Panama	Contragula	contragula	cont2	male	dried	ucsb	461.68	80.37	0.01
M_CON	Panama	Contragula	contragula	cont3	male	dried	ucsb	461.68	84.78	0.00
M_LLL	Panama	Maristella	Maristella_LLL	LLL1	male	dried	ucsb	461.13	79.82	0.00
M_LLL	Panama	Maristella	Maristella_LLL	LLL2	male	dried	ucsb	461.68	80.37	0.00
M_LLL	Panama	Maristella	Maristella_LLL	LLL3	male	dried	ucsb	461.13	78.72	0.00
M_LLL	Panama	Maristella	Maristella_LLL	LLL4	male	dried	ucsb	460.03	79.27	0.00
M_LLL	Panama	Maristella	Maristella_LLL	LLL5	male	dried	ucsb	461.13	80.92	0.00
M_MFU	Panama	Maristella	Maristella_MFU	MFU1	male	dried	ucsb	461.13	80.37	0.00
M_MFU	Panama	Maristella	Maristella_MFU	MFU2	male	dried	ucsb	461.68	80.93	0.00
M_MFU	Panama	Maristella	Maristella_MFU	MFU3	male	dried	ucsb	461.13	80.37	0.00
M_MFU	Panama	Maristella	Maristella_MFU	MFU4	male	dried	ucsb	461.68	80.38	0.00
M_MFU	Panama	Maristella	Maristella_MFU	MFU5	male	dried	ucsb	462.23	80.38	0.00
P_ann	Belize	Photeros	Photeros_annecohenae	Pann1	unknown	live	ucsb	466.51	85.25	0.00
P_ann	Belize	Photeros	Photeros_annecohenae	Pann2	unknown	live	ucsb	469.81	85.81	0.01
P_ann	Belize	Photeros	Photeros_annecohenae	Pann3	unknown	live	ucsb	469.26	85.25	0.00
P_mor	Belize	Photeros	Photeros_morini	Pmor1	male	live	ucsb	464.32	79.72	0.05
P_mor	Belize	Photeros	Photeros_morini	Pmor2	male	live	ucsb	467.06	81.37	0.00
P_mor	Belize	Photeros	Photeros_morini	Pmor3	male	live	ucsb	467.61	83.58	0.04
P_mor	Belize	Photeros	Photeros_morini	Pmor4	male	live	ucsb	465.96	82.49	0.00
P_mor	Belize	Photeros	Photeros_morini	Pmor5	male	live	ucsb	465.96	80.82	0.00
M_SMU	Panama	Maristella	Maristella_SMU	SMU1	male	dried	ucsb	461.13	79.82	0.00
M_SMU	Panama	Maristella	Maristella_SMU	SMU2	male	dried	ucsb	460.58	79.27	0.00
M_SMU	Panama	Maristella	Maristella_SMU	SMU3	male	dried	ucsb	460.58	79.82	0.00

M_SMU	Panama	Maristella	Maristella_SMU	SMU4	male	dried	ucsb	461.68	80.37	0.00
M_SMU	Panama	Maristella	Maristella_SMU	SMU5	male	dried	ucsb	460.58	79.27	0.00
K_has	Belize	Kornickeria	Kornickeria_hastingsi	Khas1	male	dried	ucsb	464.32	80.74	0.11
K_has	Belize	Kornickeria	Kornickeria_hastingsi	Khas2	male	live	ucsb	471.46	75.78	0.09
K_has	Belize	Kornickeria	Kornickeria_hastingsi	Khas3	male	live	ucsb	460.48	77.98	0.00
M_chi	Belize	Maristella	Maristella_chicoi	MSH1	male	live	ucsb	461.57	79.10	0.00
M_chi	Belize	Maristella	Maristella_chicoi	MSH2	male	live	ucsb	461.57	80.21	0.00
M_chi	Belize	Maristella	Maristella_chicoi	MSH3	male	live	ucsb	458.83	79.10	0.02
M_chi	Belize	Maristella	Maristella_chicoi	MSH4	male	live	ucsb	460.48	79.10	0.00
M_chi	Belize	Maristella	Maristella_chicoi	MSH5	male	live	ucsb	461.57	78.55	0.00
M_SVU	Belize	Maristella	Maristella_SVD	SVD1	male	live	ucsb	461.02	79.66	0.00
M_SVU	Belize	Maristella	Maristella_SVD	SVD2	male	live	ucsb	462.12	82.42	0.00
M_SVU	Belize	Maristella	Maristella_SVD	SVD3	male	live	ucsb	459.93	79.10	0.00
M_SVU	Belize	Maristella	Maristella_SVD	SVD4	male	live	ucsb	461.57	79.65	0.00
M_SVU	Belize	Maristella	Maristella_SVD	SVD5	male	live	ucsb	461.02	78.55	0.00
M_SVU	Belize	Maristella	Maristella_SVU	SVU1	male	live	ucsb	461.02	79.10	0.00
M_SVU	Belize	Maristella	Maristella_SVU	SVU2	male	live	ucsb	462.12	79.66	0.00
M_SVU	Belize	Maristella	Maristella_SVU	SVU3	male	live	ucsb	459.93	77.99	0.00
M_SVU	Belize	Maristella	Maristella_SVU	SVU4	male	live	ucsb	460.48	79.10	0.00
M_SVU	Belize	Maristella	Maristella_SVU	SVU5	male	live	ucsb	459.93	79.10	0.00
P_GPH	Roatan	Photeros	Photeros_GPH	GPH1	male	live	ucsb	466.19	75.75	0.01
P_GPH	Roatan	Photeros	Photeros_GPH	GPH2	male	live	ucsb	466.19	75.75	0.08
P_GPH	Roatan	Photeros	Photeros_GPH	GPH3	male	live	ucsb	465.63	76.30	0.01
P_GPH	Roatan	Photeros	Photeros_GPH	GPH4	male	live	ucsb	467.84	75.21	0.06
P_GPH	Roatan	Photeros	Photeros_GPH	GPH5	male	live	ucsb	466.74	77.42	0.00
P_GPH	Roatan	Photeros	Photeros_GPH	GPH6	male	live	ucsb	465.63	77.41	0.00
P_GPH	Roatan	Photeros	Photeros_GPH	GPH7	male	live	ucsb	464.53	76.31	0.00
P_GPH	Roatan	Photeros	Photeros_GPH	GPH8	male	live	ucsb	465.08	77.42	0.01
P_WLU	Roatan	Photeros	Photeros_WLU	WLU1	male	live	ucsb	466.19	79.08	0.00
P_WLU	Roatan	Photeros	Photeros_WLU	WLU2	male	live	ucsb	463.98	77.41	0.00
P_WLU	Roatan	Photeros	Photeros_WLU	WLU3	male	live	ucsb	466.19	78.52	0.00
P_WLU	Roatan	Photeros	Photeros_WLU	WLU4	male	live	ucsb	464.53	78.52	0.00
P_WLU	Roatan	Photeros	Photeros_WLU	WLU5	male	live	ucsb	466.74	80.19	0.00
P_WLU	Roatan	Photeros	Photeros_WLU	WLU6	male	live	ucsb	465.63	77.97	0.00
P_WLU	Roatan	Photeros	Photeros_WLU	WLU7	male	dried	ucsb	466.74	78.53	0.03
P_WLU	Roatan	Photeros	Photeros_WLU	WLU8	male	dried	ucsb	468.40	81.87	0.01
M_LSD	PR_USA	Maristella	Maristella_LSD	LSD1	male	live	ucsb	459.38	80.19	0.02
M_LSD	PR_USA	Maristella	Maristella_LSD	LSD2	male	live	ucsb	459.93	79.08	0.03
M_IR	Roatan	Maristella	Maristella_IR	IR1	male	dead	ucsb	459.56	75.71	0.00

M_IR	Roatan	Maristella	Maristella_IR	IR2	male	dead	ucsb	459.56	76.82	0.00
M_IR	Roatan	Maristella	Maristella_IR	IR3	male	live	ucsb	466.74	83.46	0.13
M_IR	Roatan	Maristella	Maristella_IR	IR4	male	dead	ucsb	459.56	74.60	0.01
M_IR	Roatan	Maristella	Maristella_IR	IR5	male	dead	ucsb	462.32	77.93	0.01
M_IR	Roatan	Maristella	Maristella_IR	IR6	male	dead	ucsb	462.32	75.71	0.01
M_RD	Roatan	Maristella	Maristella_RD	RD1	male	dead	ucsb	463.42	75.71	0.02
M_RD	Roatan	Maristella	Maristella_RD	RD2	male	dead	ucsb	459.56	77.38	0.01
M_RD	Roatan	Maristella	Maristella_RD	RD3	male	dead	ucsb	466.74	77.92	0.10
M_RD	Roatan	Maristella	Maristella_RD	RD4	male	dead	ucsb	465.08	78.48	0.03
M_RD	Roatan	Maristella	Maristella_RD	RD5	male	dead	ucsb	465.63	75.16	0.04
M_DU	Roatan	Maristella	Maristella_DU	DU1	male	live	ucsb	460.11	75.16	0.00
M_DU	Roatan	Maristella	Maristella_DU	DU2	male	live	ucsb	459.56	75.16	0.00
M_DU	Roatan	Maristella	Maristella_DU	DU3	male	live	ucsb	457.35	75.71	0.00
M_DU	Roatan	Maristella	Maristella_DU	DU4	male	live	ucsb	458.45	73.49	0.01
M_DU	Roatan	Maristella	Maristella_DU	DU5	male	live	ucsb	459.56	75.71	0.00
V_tsu	CA_USA	Fred	Vargula_tsujii	Vtsu1	female	live	ucsb	462.12	74.14	0.01
V_tsu	CA_USA	Fred	Vargula_tsujii	Vtsu2	female	live	ucsb	461.02	74.70	0.00
V_tsu	CA_USA	Fred	Vargula_tsujii	Vtsu3	female	live	ucsb	462.12	74.14	0.01
V_tsu	CA_USA	Fred	Vargula_tsujii	Vtsu4	male	live	ucsb	458.83	74.14	0.00
V_tsu	CA_USA	Fred	Vargula_tsujii	Vtsu5	male	live	ucsb	459.93	74.14	0.00
V_tsu	CA_USA	Fred	Vargula_tsujii	Vtsu6	male	live	ucsb	459.93	73.03	0.01
V_tsu	CA_USA	Fred	Vargula_tsujii	Vtsu7	male	live	ucsb	459.93	75.25	0.00
V_tsu	CA_USA	Fred	Vargula_tsujii	Vtsu8	male	live	ucsb	461.02	74.70	0.01
K_WCU	PR_USA	Kornickeria	Kornickeria_WCU	WCU1	male	live	ucsb	462.12	81.78	0.14
K_WCU	PR_USA	Kornickeria	Kornickeria_WCU	WCU2	male	live	ucsb	458.83	79.07	0.03
K_WCU	PR_USA	Kornickeria	Kornickeria_WCU	WCU3	male	live	ucsb	463.22	82.36	0.08
K_WCU	PR_USA	Kornickeria	Kornickeria_WCU	WCU5	male	live	ucsb	450.06	74.09	0.08
K_SPU	PR_USA	Kornickeria	Kornickeria_SPU	SPU1	male	live	ucsb	457.19	76.89	0.02
K_SPU	PR_USA	Kornickeria	Kornickeria_SPU	SPU2	male	live	ucsb	458.28	76.87	0.00
K_SPU	PR_USA	Kornickeria	Kornickeria_SPU	SPU3	male	live	ucsb	461.57	78.54	0.01
K_SPU	PR_USA	Kornickeria	Kornickeria_SPU	SPU4	male	live	ucsb	454.99	73.56	0.07
K_SPU	PR_USA	Kornickeria	Kornickeria_SPU	SPU5	male	live	ucsb	460.48	81.85	0.08
K_SPU	PR_USA	Kornickeria	Kornickeria_SPU	SPU6	male	live	ucsb	461.02	77.97	0.01
K_SPU	PR_USA	Kornickeria	Kornickeria_SPU	SPU7	male	live	ucsb	451.16	71.31	0.13
V_hil	Japan	Vargula	Vargula_hilgendorfii	Vhil0123171	unknown	dried	ucsb	460.48	76.88	0.01
V_hil	Japan	Vargula	Vargula_hilgendorfii	Vhil0123172	unknown	dried	ucsb	459.38	77.98	0.00
V_hil	Japan	Vargula	Vargula_hilgendorfii	Vhil0123173	unknown	dried	ucsb	459.38	74.68	0.04

V_hil	Japan	Vargula	Vargula_hilgendorfii	Vhil0907161	unknown	dried	ucsb	459.66	80.87	0.03
V_hil	Japan	Vargula	Vargula_hilgendorfii	Vhil0907162	unknown	dried	ucsb	460.21	77.54	0.06
V_hil	Japan	Vargula	Vargula_hilgendorfii	Vhil0907163	unknown	dried	ucsb	460.21	80.87	0.03
V_hil	Japan	Vargula	Vargula_hilgendorfii	Vhil0907164	unknown	dried	ucsb	456.33	78.10	0.02
V_hil	Japan	Vargula	Vargula_hilgendorfii	Vhil0907165	unknown	dried	ucsb	456.33	76.99	0.01
V_hil	Japan	Vargula	Vargula_hilgendorfii	Vhil0907166	unknown	dried	ucsb	460.76	79.76	0.04
V_hil	Japan	Vargula	Vargula_hilgendorfii	Vhil100520161	unknown	dried	ucsb	458.28	81.85	0.01
V_hil	Japan	Vargula	Vargula_hilgendorfii	Vhil0920161	unknown	dried	ucsb	465.19	73.11	0.02
V_hil	Japan	Vargula	Vargula_hilgendorfii	Vhil0920162	unknown	dried	ucsb	465.19	72.01	0.02
V_hil	Japan	Vargula	Vargula_hilgendorfii	Vhil0920163	unknown	dried	ucsb	461.32	72.01	0.03
V_hil	Japan	Vargula	Vargula_hilgendorfii	Vhil_Japan	unknown	live	Japan	461.14	77.87	0.00
P_gra_p	Published	Photeros	Photeros_gramminicola	Pgra_huvard	unknown	unknown	published	471.14	78.58	0.32
V_hil_p	Published	Vargula	Vargula_hilgendorfii	Vhil_tsuji	unknown	unknown	published	460.54	83.11	0.33
C_noc_p	Published	Cypridina	Cypridina_noctiluca	Cnoc_ohmiya	unknown	unknown	published	454.27	79.17	0.06

Table S5 - Collection localities, size measurements, and accession numbers for vouchers for specimens used in this study

Inferred genus	Species or field code	Locality	Latitude	Longitude	Carapace length	Carapace height	Carapace ratio	BioProject	Luciferase Genbank Accession	SRA Accession	RNA Isolation	Library Preparation	Sequencing Instrument
Photeros	EGD	Bocas del Toro, Panama	9.331385	-82.25434	1.625	1.044	1.56		N/A	N/A			
Photeros	SFM	Bocas del Toro, Panama	9.331326	-82.25327	1.703	1.077	1.58		N/A	N/A			
Photeros	annecohenae	Southwater Caye, Belize	16.8116	-88.08243	1.623	1.02	1.59	PRJNA589015		SRR10860880	Qiagen Rneasy Fibrous Tissue Mini Kit	Illumina TruSeq v3	Illumina HiSeq 2500
Photeros	morini	Southwater Caye, Belize	16.8116	-88.08243	2.056	1.282	1.6	PRJNA589015		SRR10860877	Trizol	NEB Ultra RNA Library Prep Kit for Illumina	Illumina HiSeq 1500
Photeros	morini	Southwater Caye, Belize	16.8116	-88.08243	2.056	1.282	1.6	PRJNA589015		SRR10860876	Qiagen Rneasy	Illumina TruSeq v2	Illumina HiSeq 1500
Photeros	GPH	Roatan, Honduras	16.40272	-86.409	1.683	1.038	1.62		N/A	N/A			
Photeros	WLU	Roatan, Honduras	16.35806	-86.43291	1.978	1.222	1.62	PRJNA589015		SRR10811635	Trizol	NEBNextUltra IIRNA Library Prep Kit for Illumina	NextSeq 500
Photeros	mcelroyi	Discovery Bay, Jamaica						PRJNA589015		SRR10811638	Qiagen RNA	Illumina TruSeq v3	Illumina HiSeq 1500
Photeros	mcelroyi	Discovery Bay, Jamaica						PRJNA589015		SRR10811637	Qiagen RNA	Illumina TruSeq v3	Illumina HiSeq 1500
Photeros	mcelroyi	Discovery Bay, Jamaica						PRJNA589015		SRR10811645	Trizol	NEBNext Ultra RNA Library Prep Kit for Illumina	Illumina HiSeq 1500

Maristella	MFU	Bocas del Toro, Panama	9.331326	-82.253267	1.657	1.002	1.65		N/A	N/A			
Maristella	IR	Roatan, Honduras	16.35806	-86.43291	2.282	1.386	1.65		N/A	N/A			
Maristella	SVU (MWU)	Southwater Caye, Belize	16.8116	-88.08243	2.172	1.305	1.66	PRJNA589015		SRR1081644	Trizol	NEBNext Ultra RNA Library Prep Kit for Illumina	Illumina HiSeq 1500
Maristella	SVU (MWU)	Southwater Caye, Belize	16.8116	-88.08243	2.172	1.305	1.66	PRJNA589015		SRR1081643	Qiagen RNA	Illumina TruSeq v2	Illumina HiSeq 1000
Maristella	SVD	Southwater Caye, Belize	16.8116	-88.08243				PRJNA589015		SRR1086875	Trizol	NEB Ultra RNA Library Prep Kit for Illumina	Illumina HiSeq 1500
Maristella	SVD	Southwater Caye, Belize	16.8116	-88.08243				PRJNA589015		SRR1086874	Qiagen Rneasy	Illumina TruSeq v2	Illumina HiSeq 1500
Maristella	DU	Roatan, Honduras	16.35806	-86.43291	1.631	0.981	1.66		N/A	N/A			
Maristella	RD	Roatan, Honduras	16.358061	-86.432906	1.637	0.98	1.67		N/A	N/A			
Maristella	LLL	Bocas del Toro, Panama	9.331707	-82.255633	1.775	1.057	1.68		N/A	N/A			
Maristella	SMU	Bocas del Toro, Panama	9.331707	-82.255633	2.162	1.289	1.68		N/A	N/A			
Maristella	chicoi	Southwater Caye, Belize	16.8116	-88.08243	1.624	0.963	1.69	PRJNA589015	N/A	SRR1081640	Trizol	NEBNext Ultra RNA Library Prep Kit for Illumina	Illumina HiSeq 1500
Maristella	chicoi	Southwater Caye, Belize	16.8116	-88.08243	1.624	0.963	1.69	PRJNA589015	N/A	SRR1081639	Qiagen RNA	Illumina TruSeq v2	Illumina HiSeq 1500
Maristella	VAD	Discovery Bay, Jamaica						PRJNA589015		SRR1081642	Qiagen RNA	Illumina TruSeq v3	Illumina HiSeq 1500
Maristella	VAD	Discovery Bay, Jamaica						PRJNA589015		SRR1081641	Trizol	NEBNextUltraRNA Library Prep Kit	Illumina MiSeq

												for Illumin	
Maristella	AG	Roatan, Honduras	16.358061	-86.432906				PRJNA589015		SRR10811636	Trizol	NEBNextUltra IIRNA Library Prep Kit for Illumina	NextSeq 500
C-group	CONT	Bocas del Toro, Panama	9.331707	-82.255633	1.846	1.066	1.73		N/A	N/A			
"Vargula"	tsujii	Catalina Island, CA, USA	33.445139	-118.4845	1.581	0.913	1.73	PRJNA287212		SRR1269674			
Kornickeria	SPU	Isla Magueyes, PR, USA	17.961089	-67.052156	1.383	0.78	1.77		N/A	N/A			
Kornickeria	hastingsi carriebowae	Southwater Caye, Belize	16.8116	-88.08243	1.799	1.014	1.77	PRJNA589015		SRR10860879	Qiagen Rneasy	Unknown; amplified by Novogene Co (Davis CA)	Illumina HiSeq 1500
Kornickeria	hastingsi carriebowae	Southwater Caye, Belize	16.8116	-88.08243	1.799	1.014	1.77	PRJNA589015		SRR10860878	Qiagen Rneasy	Illumina TruSeq v2	Illumina HiSeq 1500
Maristella	LSD	Isla Magueyes, PR, USA	17.961089	-67.052156					N/A				
"Vargula"	hilgendorfii	Carolina Biological Purchase	N/A		N/A				AAA30332				
Cypridina	noctiluca	NCBI							BBG57195				

Table S6 - Luciferase-specific primers to amplify from cDNA

Primer name	Target species	Sequence	Purpose
KHC Forward	*Kornickeria hastingsi carriebowae*	GGACTCGAGAAGAGAGAGGCTAAAGATTGTTTTGAATCATCTTTCC	Amplify from cDNA
KHC Reverse	*Kornickeria hastingsi carriebowae*	AGCGGCCGCTTTGGTGCATTGAGGTGG	Amplify from cDNA
PMO Forward	*Photeros morini*	TAACTCGAGAAGAGAGAGGCTCAAGAATGCGCTCAGACA	Amplify from cDNA
PMO Reverse	*Photeros morini*	AGCGGCCGCCTGGCATATGGCTGGTAC	Amplify from cDNA
SVU Forward	SVU (undescribed)	GGGCTCGAGAAGAGAGAGGCTCAAGATTGTTATGAATTCACA	Amplify from cDNA
SVU Reverse	SVU (undescribed)	AGCGGCCGCTTTACACTGAGGGGGATA	Amplify from cDNA
SVD Forward	SVD (undescribed)	GTGCTCGAGAAGAGAGAGGCTGAAGATTGTTATGAATTCACATG	Amplify from cDNA
SVD Reverse	SVD (undescribed)	AGCGGCCGCTTTACACTGAGGTGGATAC	Amplify from cDNA
MSH Forward	*Maristella chicoi*	CGGCTCGAGAAGAGAGAGGCTGAAGATTGTTATGAATTCACAT	Amplify from cDNA
MSH Reverse	*Maristella chicoi*	AGCGGCCGCTTTACACTGAGGGGGATAC	Amplify from cDNA
VTS Forward	*Vargula tsujii*	CCGCTCGAGAAGAGAGAGGCTCAAGATTGTTATGAATCAACATG	Amplify from cDNA
VTS Reverse	*Vargula tsujii*	AGCGGCCGCTTTACACTGAGGAGGATACT	Amplify from cDNA
454-Forward-Vt	*Vargula tsujii*	CTCGAGATCAGTCGAGAAACGAAACGTGATA	Amplify from cDNA
454-Reverse-Vt	*Vargula tsujii*	GAATTCCTTTACACTGAGGGGGATACTG	Amplify from cDNA
AOX Forward	n/a	GACTGGTTCCAATTGACAAGC	Sequence from clones
AOX Reverse	n/a	GCAAATGGCATTCTGACATCC	Sequence from clones

References Cited in Supplement

1. N. M. Hensley, *et al.*, Phenotypic evolution shaped by current enzyme function in the bioluminescent courtship signals of sea fireflies. *Proceedings of the Royal Society B* **286**, 20182621 (2019).

2. N. J. Reda, *et al.*, Maristella, a new bioluminescent ostracod genus in the Myodocopida (Cypridinidae). *Zool. J. Linn. Soc.* (2019) https:/doi.org/10.1093/zoolinnean/zlz082 (November 8, 2019).

3. J. D. Bendtsen, H. Nielsen, G. von Heijne, S. Brunak, Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783–795 (2004).

4. F. I. Tsuji, B. G. Nafpaktitis, T. Goto, M. J. Cormier, Spectral characteristics of the bioluminescence induced in the marine fish, Porichthys notatus, byCypridina (ostracod) luciferin. *Molecular and cellular* (1975).

5. E. A. Widder, M. I. Latz, J. F. Case, MARINE BIOLUMINESCENCE SPECTRA MEASURED

WITH AN OPTICAL MULTICHANNEL DETECTION SYSTEM. *Biol. Bull.* **165**, 791–810 (1983).

6. Y. Nakajima, K. Kobayashi, K. Yamagishi, T. Enomoto, Y. Ohmiya, cDNA cloning and characterization of a secreted luciferase from the luminous Japanese ostracod, Cypridina noctiluca. *Biosci. Biotechnol. Biochem.* **68**, 565–570 (2004).

7. K. Kawasaki, Y. Morita, S. Ohgiya, Y. Ohmiya, Y. Ohyama, Mutant luciferase. *US Patent* (2012) (August 29, 2016).

8. A. L. Huvard, Analysis of visual pigment absorbance and luminescence emission spectra in marine ostracodes (Crustacea: Ostracoda). *Comp. Biochem. Physiol. A Physiol.* **104**, 333–338 (1993).