Molecular evolution of luciferase diversified bioluminescent signals in sea fireflies (Crustacea:Ostracoda:Cypridinidae)

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

Abstract
Integrative studies of courtship signals provide rich opportunities to understand how biological diversity originates, but understanding the underlying genetics is challenging. Here we show molecular evolution of a single gene - luciferase - influenced diversification of bioluminescent signals in cypridinid ostracods, including their radiation into dozens of Caribbean species with distinctive courtship displays. We report emission spectra from twenty-one species, thirteen luciferases from transcriptomes, and *in vitro* assays that confirm function from four exemplar species. We found most sites in luciferase evolved neutrally or under purifying selection, including multiple sites that affect the color of bioluminescence in mutagenesis studies. Twelve sites in cypridinid luciferase evolved under episodic diversifying selection, including five that correlate with changes in kinetics and/or color of bioluminescence, phenotypes that manifest at the organismal level. These results demonstrate how neutral, pleiotropic, and/or selective forces may act on even a single gene and contribute to diversification of phenotypes.

Introduction

The stunningly conspicuous diversity of courtship signals is often correlated with radiations of species (Ellis and Oakley, 2016; Mendelson and Shaw, 2005; Ritchie, 2007), sometimes called 'sexual' (Martin and Richards, 2019) or 'aesthetic' radiations (Prum, 2017). Well-known examples of aesthetic radiations include terrestrial firefly beetles (e.g. Lall et al., 1980; Lewis, 2016), birds of paradise (Ligon et al., 2018; Prum, 2012), and jumping spiders (Girard and Endler, 2014; Masta and Maddison, 2002). Like many traits, courtship signals are shaped by sexual and natural selection, non-adaptive (stochastic) factors, and developmental and historical constraints (Seilacher, 1970; Tinbergen, 1963). Disentangling these factors and identifying links between DNA sequences, phenotypes, fitness, and the phylogenetic history of signals will go a long way toward clarifying how biodiversity originates and evolves, as evidenced by research programs on organismal groups such as orchid bees (Brand et al., 2020, 2015), mormyrid fishes (Nagel et al., 2017; Zakon et al., 2008, 2006), and Hawaiian swordtail crickets (Xu and Shaw, 2019). Despite recent progress, understanding the genetic causes of diversification of signals remains challenging because behavioral phenotypes result from complex processes at many levels of biological organization.

Bioluminescent ostracods of the family Cypridinidae provide an attractive system for integrative understanding of signal diversification (Gerrish and Morin, 2016; Hensley et al., 2019; Morin, 2019; Rivers and Morin, 2008). Bioluminescent cypridinids live throughout the world and use light as an anti-predator display, including a subset of species restricted geographically to the Caribbean that also use light for courtship. This Caribbean clade diversified into dozens of species, each with a courtship display that seems to be conserved within, but variable between species (Cohen and Morin, 2010; Morin, 2019). Although both sexes use light for anti-predator displays by expelling large amounts of light-producing chemicals mixed with mucus, as far as we know, only males of Caribbean species produce courtship signals. In contrast to the large cloud of light produced during anti-predator signals, the courtship signals are comprised of delicate, discrete pulses of light, formed from secreted mucus as males rapidly swim (Morin, 1986; Rivers and Morin, 2008). Multiple light-courting species commonly live in geographical sympatry, yet produce signals above different specialized microhabitats. The courtship signals also vary in other parameters, including orientation (horizontal/vertical), direction (up/down/angled), specific time of onset (Gerrish et al., 2009), and time and distance between light pulses (Gerrish and Morin, 2016). The pulses themselves vary in brightness, kinetics, and color (Harvey, 1924; Hensley et al., 2019; Rivers and Morin, 2008). Once ostracods secrete the bioluminescent signals, the phenotype of the signal 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 the same across all ostracod species in this family (Harvey, 1924), 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.

Because c-luciferases may also be expressed in vitro to test biochemical functions (Tochigi et al., 2010), this system may fit squarely into the “functional synthesis” (Dean and Thornton, 2007), 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 (Fallon et al., 2018; Viviani et al., 2011), previous studies that characterize luciferase genes of cypridinid bioluminescence exist for only two species (Nakajima et al., 2004; Thompson et al., 1989). 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.

The rate of decay of light production is one target for connecting genotype, phenotype, and diversification in cypridinids. Recent work shows decay rates vary considerably across species of luminous ostracods and are correlated non-linearly with the temporal duration of light production in individual pulses of courtship signals (Hensley et al., 2019). Hensley et al. (2019) measured kinetics of light production by fitting parameters describing exponential decay for each species’ bioluminescence. Cypridinid bioluminescence is a single-order biochemical reaction whose rate depends only on the concentration of substrate. Even though other components within the mucus could alter this first order reaction, light production from the first order reaction fades exponentially once substrate becomes limiting. Therefore, both the amount of substrate in a pulse and decay parameters of the light reaction dictate the duration of a bioluminescent pulse (Hensley et al., 2019). Although the amount of substrate secreted in a pulse is a behavioral phenotype that may be difficult to connect to genetics, enzyme kinetics - including rate of decay of light production - depends on c-luciferase proteins. We hypothesize that differences in sequences of c-luciferase enzymes influence differences in the duration of anti-predation clouds and courtship pulses, which vary dramatically across species. The duration of anti-predator light could be under natural selection and the duration of courtship pulses could be under sexual selection, used for species recognition, and/or used for choosing between males of the same species.

In addition to decay kinetics, the color of cypridinid bioluminescence could be dictated by differences in c-luciferase proteins. However unlike decay kinetics, emission spectra (“colors”) are not well-characterized in many cypridinids. Previously published experiments hinted at variation in color of ostracod bioluminescence (Harvey, 1924; Huvard, 1993), yet understanding differences between species was difficult because most spectra were determined by different researchers with different equipment and without replication, making statistical comparisons impossible (Table S1). Harvey (1924) first noticed a difference in color between two species when he cross-reacted crude preparations of luciferase and luciferin from *V. hilgendorfii* and an unknown Jamaican species. Harvey noted the luciferase preparation of *V. hilgendorfii* catalyzed a “bluish” light and the Jamaican luciferase a “yellowish” light. He also concluded the protein (luciferase) dictates the color. More recent studies also hint at color differences because two *Photeros* species (Huvard, 1993) differ from two *Vargula* and one *Cypridina* species (Nakajima

et al., 2004; Tsuji et al., 1974; Widder et al., 1983) in wavelengths of peak emission (λ_{max}). The qualitative observations of Harvey nearly 100 years ago are consistent with the more recently published emission spectra of ostracods 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, the color of bioluminescence 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.

Previously published data on ostracods and other species suggests we will be able to connect specific mutations in luciferases to differences in color and kinetics of bioluminescence. The luciferase gene for cypridinid bioluminescence is characterized for only two ostracod species, *Cypridina noctiluca* and *Vargula hilgendorfii*, whose luciferases were sequenced and expressed *in vitro* (Nakajima et al., 2004; Thompson et al., 1989). 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 (Oakley, 2005). In addition, Kawasaki et al. (2012) reported artificial mutations in *Cypridina noctiluca* luciferase, such that when expressed *in vitro* produced emission spectra that varied in λ_{max} from 435 - 466 nm. In addition to ostracods, other bioluminescent species from diverse taxa - many of which evolved luciferases and bioluminescence independently of each other (Davis et al., 2016; Haddock et al., 2010; Widder, 2010) - have been targets of genetic studies connecting genotype and phenotype. In the cnidarian *Renilla reniformis*, multiple luciferase mutations altered emission spectra, as researchers looked for red-shifted luciferases for biotechnology applications (Loening et al., 2007). Furthermore, in terrestrial firefly beetles, one amino acid substitution in luciferase alters λ_{max} from yellow-green to red (Arslan et al., 1997). In click beetles, different cDNA alleles are expressed in light organs that produce different colors (Stolz et al., 2003). Not only color, but also kinetics are dictated by known mutations in luciferases. As one example, the luciferase of a copepod - *Gaussia* - shows a variable duration of bioluminescence in mutagenesis experiments (Welsh et al., 2009). These genetic studies illustrate how organismal phenotypes (including emission color and light kinetics), which are often important in bioluminescent signals, are linked to specific genetic mutations in luciferases. Here we characterize new luciferase genes from ostracod crustaceans and test hypotheses that molecular evolution of c-luciferases was involved in diversification of color and kinetic parameters of bioluminescent signals.

Results

Thirteen new putative luciferases. We identified 13 new genes that are 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). The putative luciferases bear multiple hallmarks of c-luciferase, including two Von Willebrand Factor-D (VWD) domains, a signal peptide, and conserved cysteine sites (Nakajima et al., 2004; Oakley, 2005; Thompson et al., 1989).

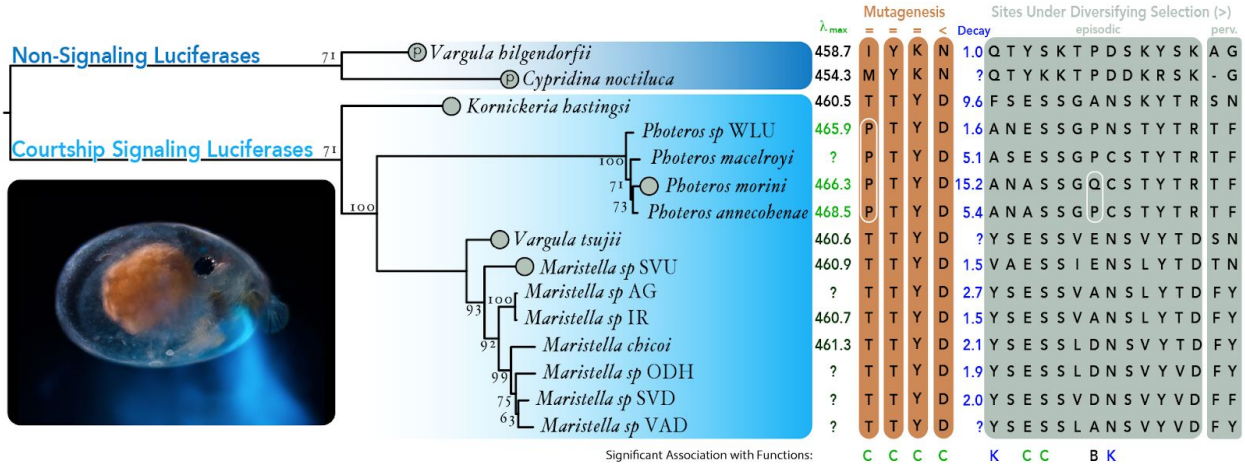

Figure 1. Maximum likelihood phylogeny of new and published c-luciferase genes using codon-aligned DNA and GTR model with fast bootstraps (percentages at nodes) in IQ-Tree. Circles at tips of phylogeny illustrate six genes expressed in vitro that show light catalysis. Two circles with "p" inside are 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 number

Patterns of natural selection in c-luciferases. Using mixed effects maximum likelihood in hyphy (Murrell et al., 2012), 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 (Kosakovsky Pond and Frost, 2005), 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. "Pervasive" positive selection refers to 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 (Viviani and Ohmiya, 2006) 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) (Fig 2). This is consistent with the putative c-luciferases across multiple genera of Cypridinidae being functional c-luciferases.

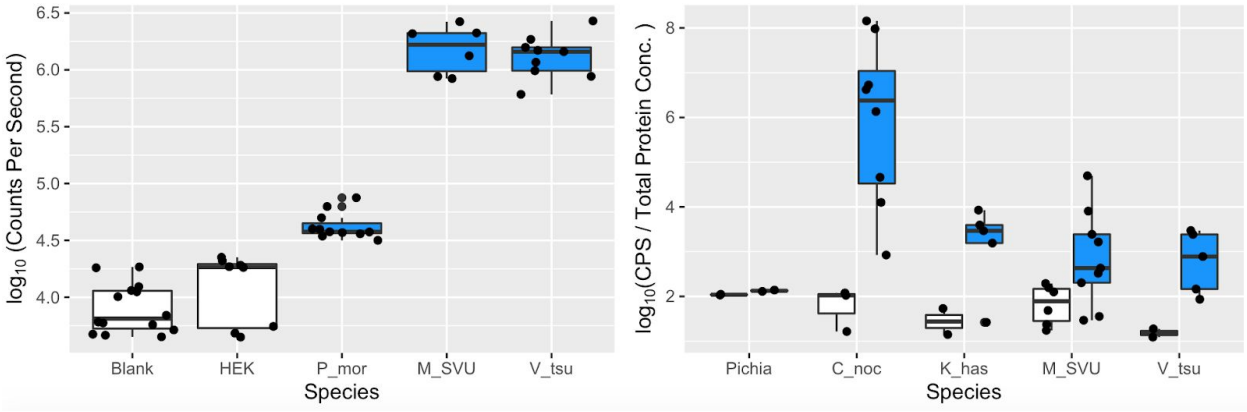

Figure 2. Exemplar luciferases are functional. **A**. Light produced after addition of luciferin to putative c-luciferase constructs from different species, measured in log Counts Per Second. Controls include blank cell culture media and HEK cells with no construct. Constructs are: (1) *Photeros morini* (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 (Cohen and Morin, 2003) and molecular (Hensley et al., 2019) phylogenetic analyses, by the phylogeny of c-luciferases we present here (Fig. 1), and by a diagnostic ratio of length:height of carapace (Gerrish and Morin, 2016; Morin and Cohen, 2017; Reda et al., 2019), 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.

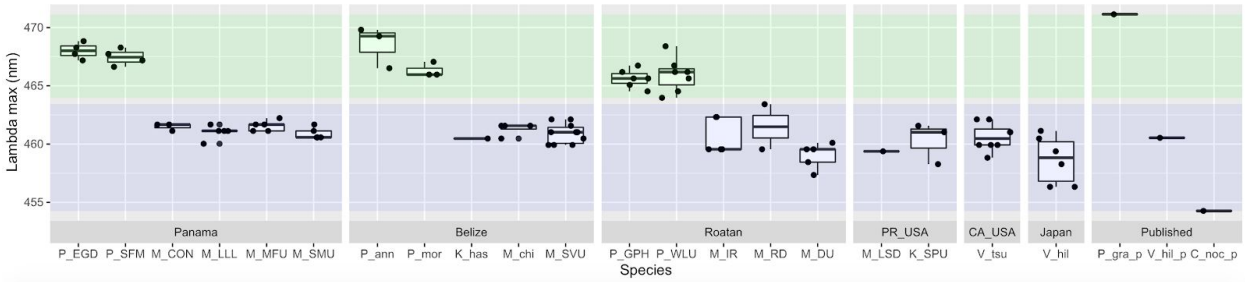

Figure 3. Wavelength value with highest emission (λ_{max}; "Lambda max" on the y-axis) in nanometers (nm) from new emission spectra from 20 species and previously published spectra from 3 species.

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 mutational analyses 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 was 435 nm and M-178-K was 447 nm, although these sequences also had mutations at other sites besides 178 (Kawasaki et al., 2012). Site 178 also varies in c-luciferases of living species, as isoleucine in *V. hilgendorfii*, threonine in *Maristella*, *Kornickeria*, and *V. tsujii*, and proline in 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 (Yokoyama et al., 2008).

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).

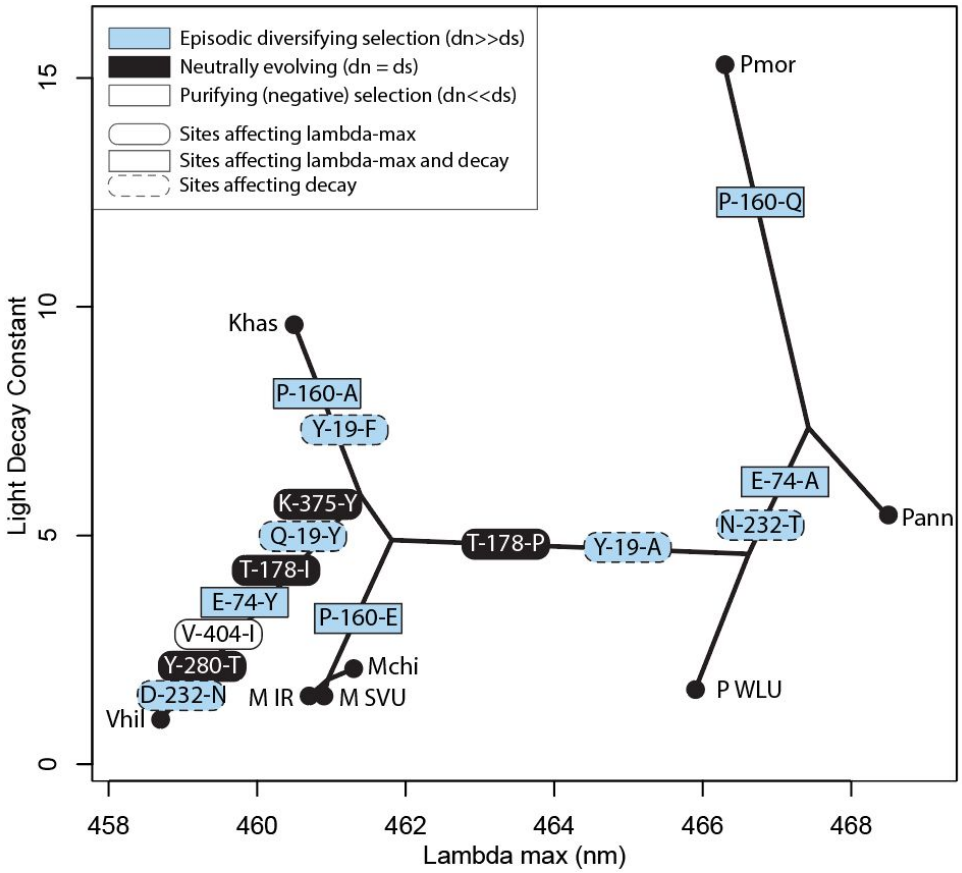

Figure 4. Bioluminescent phenotypes, including λ_{max} ("Lambda max" on the x-axis) of emission spectra and decay of light emission ("Light decay constant on the y-axis), diversified during the evolution of Cypridinidae, perhaps through different mutations in c-luciferases, and influenced by different 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

Organismal signals often vary conspicuously and may be involved in rapid radiations of species, especially when used in courtship (Ellis and Oakley, 2016; Mendelson and Shaw, 2005). The bioluminescent signals of cypridinid ostracods provide a prime example of signals that vary in multiple parameters across species. Although overall differences in signals are also strongly influenced by behavioral differences, we hypothesized some parameters of the signals rely heavily on molecular functions like enzyme kinetics and emission spectra ("color"), and therefore may be linked to specific mutations in the light-producing enzyme c-luciferase. However, only two previously published c-luciferase genes existed in the literature. Here, we report multiple new luciferase genes and biochemical phenotypes from multiple cypridinid species. Although some sites in c-luciferase that control color are evolving neutrally or under purifying selection, we also found episodic diversifying selection on luciferases at amino acid sites strongly associated with changes in both color and enzyme kinetics. These results indicate various evolutionary forces acting on a single gene may have contributed to diversification of phenotypes. In addition to connecting specific mutations to diversification of biochemical phenotypes, 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 functional and 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 to also be 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 decay parameters (Fig. 1) that are related to duration of courtship pulses (Hensley et al., 2019). 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 (Rivers and Morin, 2013). Nevertheless, we hypothesize pulse duration, which is dictated in part by enzyme kinetics

(Hensley et al., 2019), is an important parameter for fitness through inter-specific anti-predator displays and/or through species-specific mate recognition or female choice. Consistent with this hypothesis, there is extensive variation among species, with pulse durations in courtship signals ranging from less than one second in some species to more than 9 seconds in others (Cohen and Morin, 2010; Hensley et al., 2019). 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 (Nakajima et al., 2004; Yasuno et al., 2018). 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 bioluminescent decay.

Another possibility is that variation of some phenotypes of bioluminescence are not functionally relevant, 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. At the molecular level, one site (178) strongly affects λ_{max} and has a dn:ds ratio indistinguishable from one, consistent with a minimal impact on fitness. In contrast, three sites (74, 114, 160) are correlated with changes in λ_{max} and under positive selection, which could indicate selective forces actually did drive changes in color. The patterns of amino acid differences in these selected sites 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 grass (Morin, 2019), the reflectance hypothesis (Endler, 1992) 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}. 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 constant. 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 could 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 effect new mutations could have. It is also possible that site-specific epistatic interactions changed during the evolution of c-luciferases, as documented in other proteins (Ortlund et al., 2007; Yokoyama et al., 2014).

Taken together, our results illustrate the potential for varied interactions between molecular evolution, pleiotropy, epistasis, and phenotypes during diversification - even when considering a single gene like c-luciferase. We demonstrate how multiple evolutionary processes acting at a single locus may affect phenotypic evolution. When extrapolated to other genes, such as the presumably many genes affecting behavioral phenotypes and genes interacting with different environments, the combinatorics could 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 (Seilacher, 1970; Tinbergen, 1963), 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 (Cohen and Oakley, 2017). 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 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 (Hensley et al., 2019; Reda et al., 2019). We report specific collection localities and size measurements (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. 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. 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* (Oakley et al., 2012), 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 confirm nucleotide sequences with Sanger sequencing. 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. To amplify sequences from cDNA, 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. In many cases, we found complete assembly of putative luciferase genes from transcriptomes to depend on which assembly programs and parameters we used. We believe the VWD domains contained in luciferases (Oakley, 2005) 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 (Nakajima et al., 2004) 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.

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 (Bendtsen et al., 2004). 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.

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 (Haas et al., 2013), which we attribute to polymorphism from pooling individuals. In these cases, did a second assembly of luciferase fragments using cap3 (Huang and Madan, 1999) and selected the longest orfs as putative c-luciferases. We aligned these sequences using MAFFT (Katoh and Standley, 2013). We used IQTREE (Nguyen et al., 2015) 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 (Katoh and Standley, 2013), then matching DNA codons using pal2nal (Suyama et al., 2006). We used Hyphy (Pond et al., 2019) to compare ratios of synonymous to non-synonymous substitutions. We used MEME and FEL to search for individual codons under diversifying selection (Murrell et al., 2012).

Emission Spectra. We used a home-built fluorimeter (spectroradiometer) at UCSB to measure emission spectra. In earlier trials, 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. We imaged the output orifice of the integrating sphere or the 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.

We first summed all 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. For the high quality spectra, we employed Savitzky-Golay smoothing using the signal package (Ligges et al., 2015) in R (Team and Others, 2013) 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. (2012) 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 (Edgar, 2004) 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 (Hensley et al., 2019).

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 (Yokoyama et al., 2008). 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

Arslan T, Mamaev SV, Mamaeva NV, Hecht SM. 1997. Structurally Modified Firefly Luciferase. Effects of Amino Acid Substitution at Position 286. *J Am Chem Soc* **119**:10877–10887.

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

Brand P, Hinojosa-Díaz IA, Ayala R, Daigle M, Yurrita Obiols CL, Eltz T, Ramírez SR. 2020. The evolution of sexual signaling is linked to odorant receptor tuning in perfume-collecting orchid bees. *Nat Commun* **11**:244.

Brand P, Ramírez SR, Leese F, Quezada-Euan JJG, Tollrian R, Eltz T. 2015. Rapid evolution of chemosensory receptor genes in a pair of sibling species of orchid bees (Apidae: Euglossini). *BMC Evol Biol* **15**:176.

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

Cohen AC, Morin JG. 2003. Sexual Morphology, Reproduction and the Evolution of Bioluminescence in Ostracoda. *The Paleontological Society Papers* **9**:37–70.

Cohen AC, Oakley TH. 2017. Collecting and processing marine ostracods. *J Crustacean Biol*.

Davis MP, Sparks JS, Smith WL. 2016. Repeated and Widespread Evolution of Bioluminescence in Marine Fishes. *PLoS One* **11**:e0155154.

Dean AM, Thornton JW. 2007. Mechanistic approaches to the study of evolution: the functional synthesis. *Nat Rev Genet* **8**:675–688.

Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**:113.

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

Ellis EA, Oakley TH. 2016. High Rates of Species Accumulation in Animals with Bioluminescent Courtship Displays. *Curr Biol* **26**:1916–1921.

Endler JA. 1992. Signals, Signal Conditions, and the Direction of Evolution. *Am Nat* **139**:S125–S153.
Gerrish GA, Morin JG. 2016. Living in sympatry via differentiation in time, space and display characters of courtship behaviors of bioluminescent marine ostracods. *Mar Biol* **163**:190.
Gerrish GA, Morin JG, Rivers TJ, Patrawala Z. 2009. Darkness as an ecological resource: the role of light in partitioning the nocturnal niche. *Oecologia* **160**:525–536.
Girard MB, Endler JA. 2014. Peacock spiders. *Curr Biol* **24**:R588–90.
Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**:1494–1512.
Haddock SHD, Moline MA, Case JF. 2010. Bioluminescence in the sea. *Ann Rev Mar Sci* **2**:443–493.
Harvey EN. 1924. Studies on bioluminescence. XVI. What determines the color of the light of luminous animals. *Am J Physiol*.
Hensley NM, Ellis EA, Gerrish GA, Torres E, Frawley JP, Oakley TH, Rivers TJ. 2019. Phenotypic evolution shaped by current enzyme function in the bioluminescent courtship signals of sea fireflies. *Proceedings of the Royal Society B* **286**:20182621.
Huang X, Madan A. 1999. CAP3: A DNA sequence assembly program. *Genome Res* **9**:868–877.
Huvard AL. 1993. Analysis of visual pigment absorbance and luminescence emission spectra in marine ostracodes (Crustacea: Ostracoda). *Comp Biochem Physiol A Physiol* **104**:333–338.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**:772–780.
Kawasaki K, Morita Y, Ohgiya S, Ohmiya Y, Ohyama Y. 2012. Mutant luciferase. *US Patent*. 8147842.
Kosakovsky Pond SL, Frost SDW. 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* **22**:1208–1222.
Lall AB, Seliger HH, Biggley WH, Lloyd JE. 1980. Ecology of Colors of Firefly Bioluminescence. *Science*. doi:10.1126/science.210.4469.560
Lewis S. 2016. Silent Sparks: The Wondrous World of Fireflies. Princeton University Press.
Ligges U, Short T, Kienzle P, Schnackenberg S, Billinghurst D, Borchers HW, Others. 2015. signal: Signal processing. *Retrieved from Available at: https://CRAN R-project org/package= signal*.
Ligon RA, Diaz CD, Morano JL, Troscianko J, Stevens M, Moskeland A, Laman TG, Scholes E 3rd. 2018. Evolution of correlated complexity in the radically different courtship signals of birds-of-paradise. *PLoS Biol* **16**:e2006962.
Loening AM, Wu AM, Gambhir SS. 2007. Red-shifted Renilla reniformis luciferase variants for imaging in living subjects. *Nat Methods* **4**:641–643.
Martin CH, Richards EJ. 2019. The Paradox Behind the Pattern of Rapid Adaptive Radiation: How Can the Speciation Process Sustain Itself Through an Early Burst? *Annu Rev Ecol Evol Syst* **50**:569–593.
Masta SE, Maddison WP. 2002. Sexual selection driving diversification in jumping spiders. *Proc Natl Acad Sci U S A* **99**:4442–4447.
Mendelson TC, Shaw KL. 2005. Sexual behaviour: rapid speciation in an arthropod. *Nature* **433**:375–376.

Morin JG. 2019. Luminaries of the reef: The history of luminescent ostracods and their courtship displays in the Caribbean. *J Crustacean Biol* **39**:227–243.
Morin JG. 1986. Firefleas of the Sea: Luminescent Signaling in Marine Ostracode Crustaceans. *Fla Entomol* **69**:105–121.
Morin JG, Cohen AC. 2017. 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.
Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. 2012. Detecting Individual Sites Subject to Episodic Diversifying Selection. *PLoS Genetics*. doi:10.1371/journal.pgen.1002764
Nagel R, Kirschbaum F, Tiedemann R. 2017. 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.
Nakajima Y, Kobayashi K, Yamagishi K, Enomoto T, Ohmiya Y. 2004. cDNA cloning and characterization of a secreted luciferase from the luminous Japanese ostracod, Cypridina noctiluca. *Biosci Biotechnol Biochem* **68**:565–570.
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*, 179 **538**:179–192.
Oakley TH, Wolfe JM, Lindgren AR, Zaharoff AK. 2012. Phylotranscriptomics to Bring the Understudied into the Fold: Monophyletic Ostracoda, Fossil Placement and Pancrustacean Phylogeny. *Mol Biol Evol*. doi:10.1093/molbev/mss216
Ortlund EA, Bridgham JT, Redinbo MR, Thornton JW. 2007. Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* **317**:1544–1548.
Pond SLK, Kosakovsky Pond SL, Poon AFY, Velazquez R, Weaver S, Lance Hepler N, Murrell B, Shank SD, Magalis BR, Bouvier D, Nekrutenko A, Wisotsky S, Spielman SJ, Frost SDW, Muse SV. 2019. HyPhy 2.5—A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. *Molecular Biology and Evolution*. doi:10.1093/molbev/msz197
Prum RO. 2017. The Evolution of Beauty: How Darwin's Forgotten Theory of Mate Choice Shapes the Animal World - and Us. Knopf Doubleday Publishing Group.
Prum RO. 2012. Aesthetic evolution by mate choice: Darwin's really dangerous idea. *Philos Trans R Soc Lond B Biol Sci* **367**:2253–2265.
Reda NJ, Morin JG, Torres E, Cohen AC, Schawaroch V, Gerrish GA. 2019. Maristella, a new bioluminescent ostracod genus in the Myodocopida (Cypridinidae). *Zool J Linn Soc*. doi:10.1093/zoolinnean/zlz082
Ritchie MG. 2007. Sexual Selection and Speciation. *Annu Rev Ecol Evol Syst* **38**:79–102.
Rivers TJ, Morin JG. 2013. Female ostracods respond to and intercept artificial conspecific male luminescent courtship displays. *Behav Ecol* **24**:877–887.
Rivers TJ, Morin JG. 2008. Complex sexual courtship displays by luminescent male marine ostracods. *J Exp Biol* **211**:2252–2262.
Seilacher A. 1970. ARBEITSKONZEPT ZUR KONSTRUKTIONS-MORPHOLOGIE. *Lethaia*. doi:10.1111/j.1502-3931.1970.tb00830.x
Stolz U, Velez S, Wood KV, Wood M, Feder JL. 2003. Darwinian natural selection for orange bioluminescent color in a Jamaican click beetle. *Proc Natl Acad Sci U S A* **100**:14955–14959.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* **34**:W609–12.
Team RC, Others. 2013. R: A language and environment for statistical computing.
Terai Y, Seehausen O, Sasaki T, Takahashi K, Mizoiri S, Sugawara T, Sato T, Watanabe M, Konijnendijk N, Mrosso HDJ, Tachida H, Imai H, Shichida Y, Okada N. 2006. Divergent selection on opsins drives incipient speciation in Lake Victoria cichlids. *PLoS Biol* **4**:e433.
Thompson EM, Nagata S, Tsuji FI. 1989. Cloning and expression of cDNA for the luciferase from the marine ostracod Vargula hilgendorfii. *Proc Natl Acad Sci U S A* **86**:6567–6571.
Tinbergen N. 1963. On aims and methods of ethology. *Zeitschrift für tierpsychologie* **20**:410–433.
Tochigi Y, Sato N, Sahara T, Wu C, Saito S, Irie T, Fujibuchi W, Goda T, Yamaji R, Ogawa M, Ohmiya Y, Ohgiya S. 2010. Sensitive and convenient yeast reporter assay for high-throughput analysis by using a secretory luciferase from Cypridina noctiluca. *Anal Chem* **82**:5768–5776.
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.
Tsuji FI, Lynch RV, Stevens CL. 1974. Properties of luciferase from the bioluminescent crustacean, Cypridina hilgendorfii. *Biochemistry*. doi:10.1021/bi00722a024
Viviani VR, Ohmiya Y. 2006. 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.
Welsh JP, Patel KG, Manthiram K, Swartz JR. 2009. Multiply mutated Gaussia luciferases provide prolonged and intense bioluminescence. *Biochem Biophys Res Commun* **389**:563–568.
Widder EA. 2010. Bioluminescence in the ocean: origins of biological, chemical, and ecological diversity. *Science* **328**:704–708.
Widder EA, Latz MI, Case JF. 1983. MARINE BIOLUMINESCENCE SPECTRA MEASURED WITH AN OPTICAL MULTICHANNEL DETECTION SYSTEM. *Biol Bull* **165**:791–810.
Xu M, Shaw KL. 2019. The Genetics of Mating Song Evolution Underlying Rapid Speciation: Linking Quantitative Variation to Candidate Genes for Behavioral Isolation. *Genetics* **211**:1089–1104.
Yasuno R, Mitani Y, Ohmiya Y. 2018. Effects of N-Glycosylation Deletions on Cypridina Luciferase Activity. *Photochem Photobiol* **94**:338–342.
Yokoyama S, Xing J, Liu Y, Faggionato D, Altun A, Starmer WT. 2014. Epistatic adaptive evolution of human color vision. *PLoS Genet* **10**:e1004884.
Yokoyama S, Yang H, Starmer WT. 2008. Molecular basis of spectral tuning in the red- and green-sensitive (M/LWS) pigments in vertebrates. *Genetics* **179**:2037–2043.
Zakon HH, Lu Y, Zwickl DJ, Hillis DM. 2006. Sodium channel genes and the evolution of diversity in communication signals of electric fishes: convergent molecular evolution. *Proc Natl Acad Sci U S A* **103**:3675–3680.
Zakon HH, Zwickl DJ, Lu Y, Hillis DM. 2008. Molecular evolution of communication signals in electric fish. *J Exp Biol* **211**:1814–1818.

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