Transcriptome phylogenies support ancient evolutionary transitions in bioluminescence traits of cypridinid ostracods

Authors:
Emily A. Ellis[1,2*], Jessica A. Goodheart[1,3*], Nicholai M. Hensley[1,4], Vanessa L. González[5], Nicholas J. Reda[6], Trevor J. Rivers[7], James G. Morin[8], Elizabeth Torres[9], Gretchen A. Gerrish[6,10], Todd H. Oakley[1]

[*] Equal authorship
[1] Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, Santa Barbara, California 93106, USA
[2] *Current address:* Illumina, California, USA
[3] *Current address:* Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92037, USA
[4] *Current address:* Department of Neurobiology and Behavior, Cornell University, Ithaca, New York 14850, USA
[5] Global Genome Initiative, Smithsonian Institution, National Museum of Natural History, 10th and Constitution NW, Washington, DC 20560-0105, USA
[6] Biology Department, University of Wisconsin–La Crosse, La Crosse, Wisconsin 54601, USA
[7] Department of Ecology and Evolutionary Biology, University of Kansas
[8] Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York 14850
[9] Department of Biological Sciences, California State University Los Angeles
[10] Current address: Trout Lake Station, Center for Limnology, University of Wisconsin - Madison

Abstract

Bioluminescence evolved many times independently, leading to dramatic effects on ecosystems by influencing communication both within and between species. One origin of bioluminescence is within cypridinid ostracods. Bioluminescent cypridinids probably all use light as an anti-predator display, while a subset that diversified in the Caribbean also use light for courtship signaling. Despite their importance for understanding the evolution of bioluminescence, very little molecular phylogenetic data are available for cypridinids and the timing of evolutionary transitions of luminous traits is poorly understood. Here, we estimate the first transcriptome-based molecular phylogeny and divergence times of Cypridinidae. Our results strongly support previous hypotheses of a single origin of bioluminescent courtship signaling, nested within a single origin of bioluminescence, and the secondary loss of courtship signaling in *Vargula tsujii*. We propose the name Luminini for the Tribe of bioluminescent cypridinids and Luxorina for the Sub-tribe of cypridinids with courtship signaling. Our relaxed-clock estimates of divergence times coupled with stochastic character mapping show luminous courtship evolved at least 151 Million Years Ago (MYA) and cypridinid bioluminescence originated at least 197 MYA, making it one of the oldest documented origins of bioluminescence. The molecular phylogeny of cypridinids will serve as a foundation for integrative and comparative studies on the biochemistry, molecular evolution, courtship, diversification, and ecology of cypridinid bioluminescence.

Introduction

Bioluminescence evolved separately at least 94 times (Lau and Oakley 2020), often functioning in defense and reproduction to affect ecosystems (Morin 1983; Wilson and Hastings 1998; Haddock et al. 2010; Widder 2010). When used for courtship, bioluminescence is associated with more species, providing an example of a sexually selected trait that may increase rates of speciation (Ellis and Oakley 2016). Despite recent attention, the specific number and timing of origins of bioluminescence remains uncertain in the absence of well-supported phylogenetic hypotheses and explicit estimates of divergence times for many luminous groups and their relatives. In particular, analyses of bioluminescence with fossil calibrations and models of molecular clocks are rare (Haddock et al. 2010; Ellis and Oakley 2016). As one exception, a meta-analysis reported the timing of several separate origins of bioluminescence, ranging from 20 MYA in the *Archamia* cardinalfishes to 255 MYA in sepiolid squid (Ellis and Oakley 2016).

However, most of these times were derived from studies not focused specifically on the evolution of bioluminescence and so are limited by taxon sampling. Many other origins of bioluminescence do not yet have enough phylogenetic information available to reliably infer timings of transitions in traits involved in bioluminescence. Here, we report new phylotranscriptomic analyses and divergence time estimates for cypridinid ostracods, a family of crustaceans that contains at least 250 species across 30 genera (Morin 2019), and which evolved bioluminescence separately from other animals.

Approximately half of all described cypridinids are bioluminescent (Morin 2019), and these secrete defensive clouds of light in response to predation attempts (Morin and Cohen 2010; Rivers and Morin 2012), a trait observed in many cypridinid species around the world. While all bioluminescent cypridinids probably produce this anti-predatory display, about 75 described and nominal species also produce bioluminescent courtship signals (Morin 1986, 2019; Reda et al. 2019), a trait observed only among species from the Caribbean (Cohen and Morin 1990a, 1990b; Morin and Cohen 2010; Reda et al. 2019). Male cypridinids in the Caribbean produce species-specific displays as patterns of ephemeral pulses of light, secreted as they swim in tight spiral patterns (Rivers and Morin 2008). Although relatively constant within species, courtship displays vary among species in ecological contexts, especially in the time of onset and microhabitat background of the displays (Gerrish et al. 2009; Gerrish and Morin 2016). Parameters of the display itself also vary among species, including orientation (horizontal, diagonal, or vertical) and direction (up/down) of swimming, number of pulses, time and distance between pulses, and duration of pulses (Morin 1986; Gerrish and Morin 2016; Hensley et al. 2019).

The evolutionary radiation of distinct courtship displays of cypridinids is somewhat reminiscent of the courtship signals of terrestrial fireflies, with a few key differences. Cypridinid displays begin in near darkness (Gerrish et al. 2009), while many fireflies start signaling at dusk (Lall et al. 1980). Whereas variation in color of cypridinid displays is very small (Hensley et al. 2021), fireflies produce a wider variety of luminescent colors (Branchini et al. 2004). Cypridinids also have larger multi-species assemblages (6-8 species) conducting courtship displays in overlapping time and space (Gerrish and Morin 2016), whereas fireflies rarely have more than 2-3 species observed at one location and courtship behaviors are more spatially and temporally isolated. Finally, cypridinid females are not known to ‘duet’ by using their own light to call to males during courtship displays, but female terrestrial fireflies may signal to males (reviewed in Bailey 2003). Instead, cypridinid females locate a conspecific male and intercept his path to meet beyond the final courtship pulses (Rivers and Morin 2013).

Previous phylogenetic analyses of morphology (Cohen and Morin 2003) and rDNA (Torres and Gonzalez 2007) inferred a single origin of bioluminescence and suggested a single origin of luminous courtship in Cypridinidae (Morin 2019). Using 58 morphological characters from nearly all genera of cypridinids, Cohen and Morin (2003) used parsimony with different character weighting to find strong support for monophyly of cypridinids with bioluminescence. Using 16S and 12S rDNA sequences from 36 species, Torres and Gonzalez (2007) found support for a single origin of bioluminescent cypridinids, albeit with low bootstrap support for the critical node (64%). These same previous analyses also suggested a single origin of luminescent courtship, nested within the clade of bioluminescent species. Morphological analyses indicated that the most parsimonious topologies with weighted characters supported a single origin of luminous courtship but the unweighted analysis and bootstrap consensus were ambiguous (Cohen and Morin 2003). The rDNA analyses also suggested a single origin of luminous courtship in cypridinids, although with ambiguity from a low bootstrap support value of 56% for the node (Torres and Gonzalez 2007). Both morphology and rDNA found *Vargula tsujii* - a bioluminescent cypridinid from the Eastern Pacific - to nest phylogenetically within the courtship-signaling clade, even though luminous courtship is not known and probably absent in *V. tsujii (Goodheart et al. 2020)*. Beyond the topology itself, we have very little information about divergence times for these evolutionary transitions. Cohen and Morin (2003) did suggest a recent origin of luminous courtship in cypridinids because the trait is restricted to the Caribbean, which itself originated with the close of the Isthmus of Panama some 3.5-25 MYA (Coates et al. 1992; Farris et al. 2011; Bacon et al. 2015; O'Dea et al. 2016 and references therein). Additionally, one study of divergence times of Pancrustacea included two luminous ostracods, resulting in very large confidence intervals on the time of origin of their common ancestor, ranging from about 50-250 MYA (Oakley et al. 2013).

Here, we present the first transcriptomic phylogenies and divergence time estimates for Cypridinidae. We produced phylogenies using protein coding genes inferred from newly sequenced transcriptomes, which are also useful for understanding the genetics and evolution of bioluminescence (Hensley et al. 2021). Combined with high morphological (Cohen and Morin 1990b, 2003) and biochemical (Shimomura and Johnson 1971; Thompson et al. 1989) similarity of bioluminescent phenotypes, our results support a single origin of cypridinid bioluminescence and a single origin of luminous courtship. Regardless of inherent variability in relaxed molecular clock estimates of divergence times, our most conservative estimate for the origin of cypridinid bioluminescence is at least 197 MYA and our best estimate for the origin of cypridinid bioluminescence is 267 MYA making it one of the oldest well-documented origins of

bioluminescence (Ellis and Oakley 2016). We conservatively estimate the origin of bioluminescent courtship in cypridinids to be at least 151 MYA, with a favored estimate of 213 MYA, clearly and significantly pre-dating the global climate and oceanographic changes associated with the closure of the Isthmus of Panama (Baumgartner et al. 2013) that defines the origin of the Caribbean Sea where these animals live today. Despite some evidence of gene discordance, we find strong and consistent support for many genus- and species-level relationships within Cypridinidae, and we identify a clade of close Caribbean relatives of the only known luminous cypridinid in California, *Vargula tsujii*.

Materials and Methods

Sampling Strategy and Collection

Based on previous morphological phylogenies and known distributions of species, we aimed to balance cost-efficiency and diversity of sampling by choosing eight primary sampling localities. We targeted species with luminous courtship mainly from five localities: Jamaica, Honduras (Roatan), Belize, Panama, and Puerto Rico. We targeted species outside the courtship signaling clade from three localities: Australia, Japan, and the United States. We collected ostracods via net sampling during courtship displays, sediment sampling or carrion traps (see Cohen and Oakley (2017) for sampling methods and Table S1 for taxa sampled). We preserved samples in 95% ethanol for vouchers and RNAlater for sequencing. To collect taxa with luminous courtship (where many species are undescribed), we targeted putative species by identifying unique courtship displays. We measured length and height of carapaces to tentatively assign species to genera (Table S2). The ratio of carapace length to height is usually diagnostic for genera of signaling cypridinids (Gerrish and Morin 2016; Hensley et al. 2019; Reda et al. 2019).

Taxon Sampling and Sequencing

We generated new RNA-Seq data from 45 species of cypridinid ostracods and four species of non-cypridinid ostracods. We included outgroups from all four non-cypridinid families of Myodocopida, sequencing one species each of Rutidermatidae and Sarsiellidae, and two of Philomedidae. We also used previously published sequence data from one cylindroleberid (SRX2085850) (Schwentner et al. 2018), and predicted proteins from the genome of *Darwinula stevensoni* (ENA PRJEB38362) (Schön et al. 2021). For transcriptome sequencing, we

combined RNA from whole bodies of up to 10 individuals. For some samples we extracted RNA using Trizol, and for the rest we used Qiagen RNEasy Kits. We either used the Illumina or NEBNext Ultra RNA Library Prep Kits to prepare transcriptome libraries (Table S1).

Quality Control and Dataset Assembly

We trimmed RNA-seq adapters and low-quality bases (<20) using TrimGalore v0.4.1 (Krueger 2012), then assembled trimmed reads using Trinity v2.2.0 (Grabherr et al. 2011). In cases where we produced multiple transcriptomes from the same species, we combined raw, trimmed reads to create single assemblies. We used CroCo v0.1 (Simion et al. 2018) to remove lowly expressed contigs, as they could result from cross-contamination. We proceeded with 'clean' assemblies from CroCo for further analyses. We assessed completeness using BUSCO v3.0.1 (Simão et al. 2015) using the arthropod_odb9 single-copy ortholog database. We generally required transcriptomes to have at least 100 complete BUSCO genes for further analyses. However, we made exceptions for three species (*Photeros jamescasei, Jimmorinia gunnari,* and *Melavargula japonica*), to maintain these described and taxonomically important species within the analyses. We produced protein predictions using TransDecoder v3.0.1 (Haas et al. 2013) for phylogenetic analyses.

Orthology Determination

For analyses that relied on comparing orthologs (and excluding paralogs), we compared two approaches to determine orthologs. First, we used Orthofinder (Emms and Kelly 2019) coupled with phylopypruner (Thalén 2018). We grouped genes with Orthofinder 2.5.2 (Emms and Kelly 2019), using an MCL inflation parameter set to 2.1 and used Diamond as the similarity search program. This allowed us to cluster sequences based on similarity into "orthogroups" that contain a mix of orthologs and paralogs. Next, we created gene trees for the separate orthogroups by aligning each using the 'auto' method of MAFFT v7.305 (Katoh and Standley 2013) and using maximum likelihood implemented in FastTree v2.1.9 (Price et al. 2010), assuming a JTT+CAT. We used phylopypruner 1.2.3 to remove paralogs and retain orthologs. In phylopypruner, we required a minimum of 20 species for each gene family, required the minimum support value for a node to be 0.7, and used the LS pruning method to differentiate orthologs and paralogs. We refer to this resulting set of orthologs as the "ppp orthologs".

For the second approach, we used HaMStR v13.2.6, which determines orthologs based on HMM (Hidden Markov Model) searches (Ebersberger et al. 2009) and subsequent gene tree construction to determine orthologs. Once HaMStR identified candidates from our

transcriptomes for all 1012 orthologs in the arthropod core ortholog set, we aligned each ortholog family using the ‘auto’ method of MAFFT v7.305 (Katoh and Standley 2013), trimmed the alignments using ‘automated1’ of Trimal v1.2rev59 (Capella-Gutierrez et al. 2009), and removed sequences containing more than 50% gaps using PhyloTreePruner v1.0 (Kocot et al. 2013). We estimated gene trees with the remaining alignments using FastTree v2.1.9 (Price et al. 2010) assuming a WAG model. We then identified using LongBranchFinder.pl (Oakley et al. 2014) any sequences with branches longer than 3 standard deviations of the gene tree’s mean branch length, which may represent sequencing errors causing frame-shifts during translation. We pruned genes identified as long branched, except genes from outgroup taxa, which have more distantly related genes. After pruning, we again aligned each ortholog family using MAFFT and estimated gene trees using IQ-Tree v1.6.6 (Nguyen et al. 2015) before determining the final ortholog set using PhyloTreePruner (Kocot et al. 2013). To be considered an ortholog family in this data set, we required each maximally inclusive subtree to include five taxa and at least 90% bootstrap support in the gene tree. We called this set of orthologs the “hamster orthologs".

Previously Published Datasets

To gain a broader understanding of cypridinid relationships, we included single-gene mitochondrial data (12S, 16S, CO1), summarized in Table S5. Some of these sequences are reported previously in conference proceedings (Torres and Gonzalez 2007) and other studies (Ogoh and Ohmiya 2004; Wakayama and Abe 2006; Wang et al. 2019; Goodheart et al. 2020; Pham et al. 2020). We then identified these three mitochondrial genes from our transcriptomes by identifying the longest contigs with high similarity to published complete mitochondrial genomes of *V. hilgendorfii* and *V. tsujii* (Ogoh and Ohmiya 2004; Goodheart et al. 2020). We annotated mitochondrial data from transcriptomes using MITOS (Bernt et al. 2013). We aligned each mitochondrial gene using the ‘auto’ method of MAFFT v7.305 (Katoh and Standley 2013).

Phylogenetic Reconstruction

We compared a variety of coalescence and concatenation-based analyses using Maximum Likelihood and Bayesian approaches. For coalescence-based approaches, we used ASTRAL-Pro (Zhang et al. 2020), which assumes the multi-species-coalescent to estimate a species tree from multiple gene trees, including paralogs. We used the same gene trees as input for ASTRAL-Pro as we used for phylopypruner (described above). We also conducted concatenated analyses using either ppp-orthologs or hamster-orthologs with strategies for multiple sequence alignment explained above. Next, we determined the best-fit partitioning

scheme for each concatenated dataset using PartitionFinder v.2.1.1 (Lanfear et al. 2016) with RAxML (Stamatakis 2014) and rcluster (Lanfear et al. 2014) options. Using optimal partitioning schemes and models, we produced species trees in IQ-TREE v2.0.3, with additional options for -bnni and -alrt (set to 1000). We also calculated gene Concordance Factors (gCF) using IQ-TREE (Minh et al. 2020).

Divergence time estimation

We estimated divergence times of nodes in the cypridinid phylogeny using a Bayesian relaxed molecular clock approach in MrBayes (Ronquist and Huelsenbeck 2003). Because Bayesian methods are computationally demanding (Smith et al. 2018), we chose 15 orthologs to use for divergence time estimation, as follows. Starting with orthologs from phylopypruner (ppp-orthologs), we used sortadate (Smith et al. 2018) to rank all gene trees based on their consistency with the species tree inferred from all genes in ASTRAL-Pro and based on the most consistent root-to-tip branch lengths. Although ostracods have a dense fossil record, most fossil ostracods are podocopids and assigning fossils confidently to taxonomic groups is often challenging (Tinn and Oakley 2008). These challenges restricted us to two fossil constraints. The first was Myodocopida, for which we used an offset exponential prior with a minimum age of 448.8 MY and maximum of 509 MY (Oakley et al. 2013; Wolfe et al. 2016). We also used a uniform prior between 443.8 and 509 MY on the root of the tree, which is the common ancestor of Ostracoda. Along with node constraints, we used the Independent Gamma Rate (IGR) model, which relaxes the assumption of a strict molecular clock by allowing branch lengths to vary as a Brownian Motion process (Lepage et al. 2007; Zhang 2016). We used a prior of exp(10) for variation in the IGR model . For the prior on the overall rate of the molecular clock (substitution rate/site/MY), we assumed a lognormal distribution with a mean of 0.001 and standard deviation of 0.0007. For the model of amino acid substitution, we assumed fixed-rate models and used the command aamodelpr=mixed, which explores multiple fixed-rate models in proportion to their posterior probabilities. With these models and priors, we ran 500K steps of Markov Chain Monte Carlo (MCMC) in two chains, using a starting phylogeny based on parsimony.

Ancestral state reconstructions

We reconstructed ancestral states for bioluminescence (presence/absence) and bioluminescent courtship (presence/absence). We assessed fit for two models (for each character) using the corrected Akaike Information Criterion (AICc), where: (i) transition rates were equal (ER), and (ii) forward and reverse transitions were different between states (all rates different, ARD). The

ER model (AICc = 22.22017) was slightly better than the ARD model (AICc = 23.6381) for bioluminescence and for bioluminescent courtship (ER AICc = 24.47444; ARD AICc = 24.67609), so we assumed ER for both final analyses. We used the species tree based on combined mitochondrial and transcriptome data because it has the most complete taxon sampling. We produced an ultrametric version of the tree using the chronos function in the APE package (Paradis et al. 2004) in R and did ancestral state reconstruction analysis and model testing using rayDISC in corHMM (Beaulieu et al. 2013). corHMM fits a hidden rates model that treats rate classes as hidden states in a Markov process. To test for significance of the reconstruction at each node, we used proportional likelihood significance tests assuming a log likelihood difference of 2 or greater represents a significant difference (Pagel 1999).

To estimate the time of origin of cypridinid bioluminescence and luminous courtship, we conducted time-tree stochastic character mapping (ttscm) (Alexandrou et al. 2013). This approach simulates character evolution with stochastic character mapping (Huelsenbeck et al. 2003) along time-calibrated phylogenies from the MCMC search to represent the posterior distribution of trees. While ancestral state reconstructions depict character state changes only at nodes, stochastic character maps allow state-changes along branches. When branches are time-calibrated, this approach provides absolute estimates of the timing of character state changes. We used stochastic character mapping in phytools (Revell 2012) using the same character states and models as described for ancestral state reconstruction. We used 100 time-calibrated trees from the MCMC search to represent the posterior probability distribution of tree topologies and branch lengths, and summarized the timing character state changes in histograms.

Results

Assembly and orthology statistics

Our final data matrices of transcriptomes include 44 cypridinids from 15 of 25 genera, plus six outgroup taxa representing all four other myodocopid families and the previously published genome sequence of one podocopid (Schön et al. 2021). Some of these transcriptomes were used to identify luciferase genes (Hensley et al. 2021). Once assembled, the number of contiguously assembled sequences (“contigs”) per species ranged from 5044 (*Photeros jamescasei*) to 833,810 (*Vargula tsujii*) ($\bar{x}$ = 189,168; Table S3) and BUSCO v3 Complete Scores ranged from 6 (*Photeros jamescasei*) to 928 (*Darwinula stevensoni*), out of 1066 total

arthropoda_db9 orthologs ($\bar{x}$ = 555; Table S3). N50 ranged from 254 (*Photeros jamescasei*) to 56,422 (*Darwinula stevensoni*) ($\bar{x}$ = 1,596 bp; Table S3).

Evolution of cypridinid bioluminescence and luminous courtship

The phylogenetic results show strong support for a single clade containing all Cypridinidae that use bioluminescence for courtship signaling. The analysis of transcriptomic data using the multi-species coalescent supports this clade with 1.0 posterior probability (Figs. 1, S1, S2) and the concatenated analyses using ppp-orthologs or hamster-orthologs support the clade with 100% bootstrap confidence (Figs. 1, S3, S5). Gene concordance analysis shows 49.78% of genes (1111 of 2232 gene families from ppp-orthologs) support this node (Fig S4, S7). The Bayesian analysis focused on divergence time estimation (Figs. 2, S8-S10) and the combined mitochondrial and transcriptome dataset (Figs 3, S11, S12) also recovered this node with posterior probabilities of 1.0.

Our results also support a single clade containing all bioluminescent Cypridinidae. While this clade is supported in each of our analyses, the taxa available for each data set vary due to challenges with obtaining transcriptomic data from some remote species. In the transcriptome-based analyses, bioluminescent cypridinids share a common ancestor that is well supported, and contains only luminous species. The analysis of transcriptomic data using the multi-species coalescent supports this clade with 1.0 posterior probability (Figs. 1, S1) and the concatenated analyses using ppp-orthologs and hamster-orthologs support the clade with 100% bootstrap (Figs 1, S3, S5). Gene concordance analysis shows 45.29% of genes (77 of 170 gene families from ppp-orthologs) support this node (Fig S7). The Bayesian analysis focused on divergence time estimation also recovers this node with a posterior probability of 1.0 (Figs 2, S10).

The data set that combines mitochondrial and transcriptome data also supports a single clade containing all bioluminescent Cypridinidae (Fig 3). However, the difference in taxon sampling highlights an important result. Namely, four species (three luminous and one unknown) with only mitochondrial data form a rather weakly supported clade (0.6 PP) with the non-luminous *Jimmorinia*, which is the sister group of luminous taxa in the transcriptome analyses. The three luminous species in this clade include *Vargula norvegica* from the deep Atlantic Ocean and *Sheina orri* and *Vargula karamu* from Australia. *Vargula tubulata* is also in this clade, but the status of bioluminescence in the species is unknown. Ancestral state reconstruction analyses also indicate homology, significantly supporting a single origin of bioluminescence and one origin of luminous courtship within Cypridinidae (Figs. 3, S14, S15).

Genus-level relationships within Cypridinidae

Our analysis included a number of undescribed species, each of which we putatively assigned to a genus, using primarily the length to height ratio of the carapace (Table S2). Our phylogenetic reconstructions (Fig. 1) provide good support for these putative designations, further strengthening length:height ratio as diagnostic for the main genera of signaling cypridinids (Reda et al. 2019). Using these designations, both the partitioned and coalescent-based approaches (Fig. 1) recovered two previously described genera (*Kornickeria* and *Photeros*) as monophyletic, with strong support. For *Photeros*, the analysis of transcriptomic data using the multi-species coalescent supports the clade with 1.0 posterior probability (Figs. 1, S1) and the concatenated analyses using ppp-orthologs and hamter-orthologs support the clade with 100% bootstrap (Figs 1, S3, S5). Gene concordance analysis shows that 83.29% of genes (1730 of 2077 gene families from ppp-orthologs) support this node (Fig S7). The Bayesian analysis focused on divergence time estimation (Figs 2, S10) and the combined mitochondrial and transcriptome dataset (Fig 3) also recovered this *Photeros* monophyly with posterior probabilities of 1.0.

For *Kornickeria*, the analysis of transcriptomic data using the multi-species coalescent supports the clade with 1.0 posterior probability (Figs. 1, S1) and the concatenated analyses using ppp-orthologs or hamster-orthologs support the clade with 100% bootstrap confidence (Figs 1, S3, S5). Gene concordance analysis shows that 65.89% of genes (1219 of 1850 gene families from ppp-orthologs) support this node (Fig S7). The Bayesian analysis focused on divergence time estimation (Figs 2, S10) and the combined mitochondrial and transcriptome dataset (Fig 3) also recovered *Kornickeria* monophyly with posterior probabilities of 1.0.

A third described genus with courtship signaling is *Maristella* (Reda et al. 2019), which includes previously nominal clades, the small-bodied H-Group and the larger-bodied U-Group (Cohen and Morin 1989). *Maristella* is monophyletic, based on transcriptome data analyzed with the multi-species coalescent, which supports the clade with 1.0 posterior probability (Figs. 1, S1). The concatenated analyses support the clade with 100% bootstrap confidence (Figs 1, S3, S5). Gene concordance analysis shows that 57.34% of genes (1226 of 2138 gene families from ppp-orthologs) support this node (Fig S4, S7, Table S4). The Bayesian analysis focused on divergence time estimation (Figs 2, S10) also recovered this *Maristella* monophyly with posterior probabilities of 1.0. Although strongly supported by transcriptome data, two species with only mitochondrial data are closely related to *Maristella*, even though they may belong to a different (undescribed) genus. These include *Vargula contragula* and an undescribed species from Florida from the provisional C-group (Cohen and Morin 1986; Morin and Cohen 2017). Our

results strongly support a clade that includes both C-group and *Maristella* with high probability (1.0), but the position of the two C-group species compared to *Maristella* is unresolved (Fig 3). Our results also support monophyly of the larger-bodied U-group (BZ-SVU, RO-AG, and RO-IR), which could be a new genus distinct from the smaller-bodied species that would remain in *Maristella* (Fig 1).

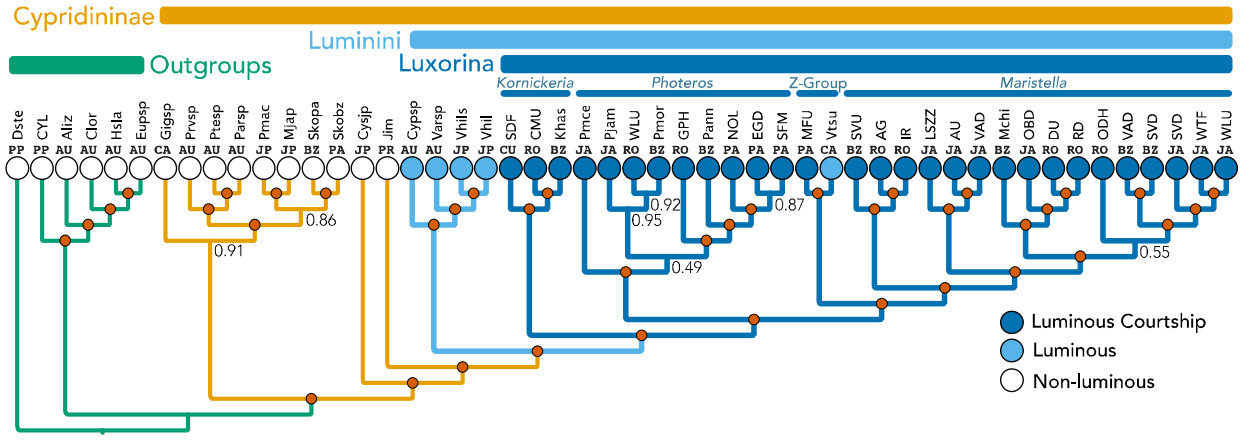


Figure 1. Species phylogeny inferred using multi-species-coalescent, employed with ASTRAL-Pro (Zhang et al. 2020) with 8485 gene trees, each of which included both orthologs and paralogs. We calculated each gene tree using fasttree 2.1.10 (Price et al. 2010) assuming a JTT+CAT model. Red circles at nodes indicate 1.0 Posterior Probability, and nodes with values less than 1.0 are indicated numerically. Circles at tips represent non-luminous (white), luminous but non-signaling (light blue), and species with luminous courtship (dark blue). The three described genera of signaling species are labeled at the top, along with "Z-group", which is a clade of signaling species plus *Vargula tsujii. Vargula* is polytypic (see also (Cohen and Morin 2003; Morin 2019)) and the type species is outside this group, so formally, these should not be *Vargula*. Above the tree, we also labeled non-cypridinid Outgroups (green), and Cypridinidinae (Orange). We herein suggest a name for the bioluminescent Cypridinidae to be Tribe Luminini (for lumen = light and the typical Tribe suffix for zoology) and a name for the clade containing signaling cypridinids to be Sub-Tribe Luxorina (lux = light + uxoriae ~= courtship + -ini = typical zoological subtribe suffix). Locality abbreviations for collecting sites of species are above each tip as follows: AU = Australia, BZ=Belize, CA=California, USA, CU=Curacao, JA= Jamaica, JP=Japan, PA=Panama, PP=Previously Published, PR = Puerto Rico, and RO=Roatan, Honduras (see Table S1 for specific localities). Above locality names are abbreviations for species. Described species are abbreviated as the first letter of genus and first three letters of species epithet; undescribed species outside of Luxorina are labeled as first three letters of genus followed by sp or two locality or collector letters when there is more than one undescribed species from the genus; within Luxorina, we used 2-4 letter field codes for undescribed species (see supplement for tree with full species names).

Another well-supported clade includes representatives of what is provisionally called "Z-group" (Cohen and Morin 1986; Morin and Cohen 1988, 2017), which will likely need description as a new genus. In the analysis of mitochondrial and transcriptome data, this clade contains *Vargula kuna*, *V. mizonomma*, and the undescribed species MFU, all from Panama; plus *Vargula tsujii* from California. *Vargula kuna* and *V. mizonomma* from San Blas do not have

transcriptome data but the two species with transcriptome data group together with strong support. Our results support *Vargula tsujii* as nested within "Z-group".

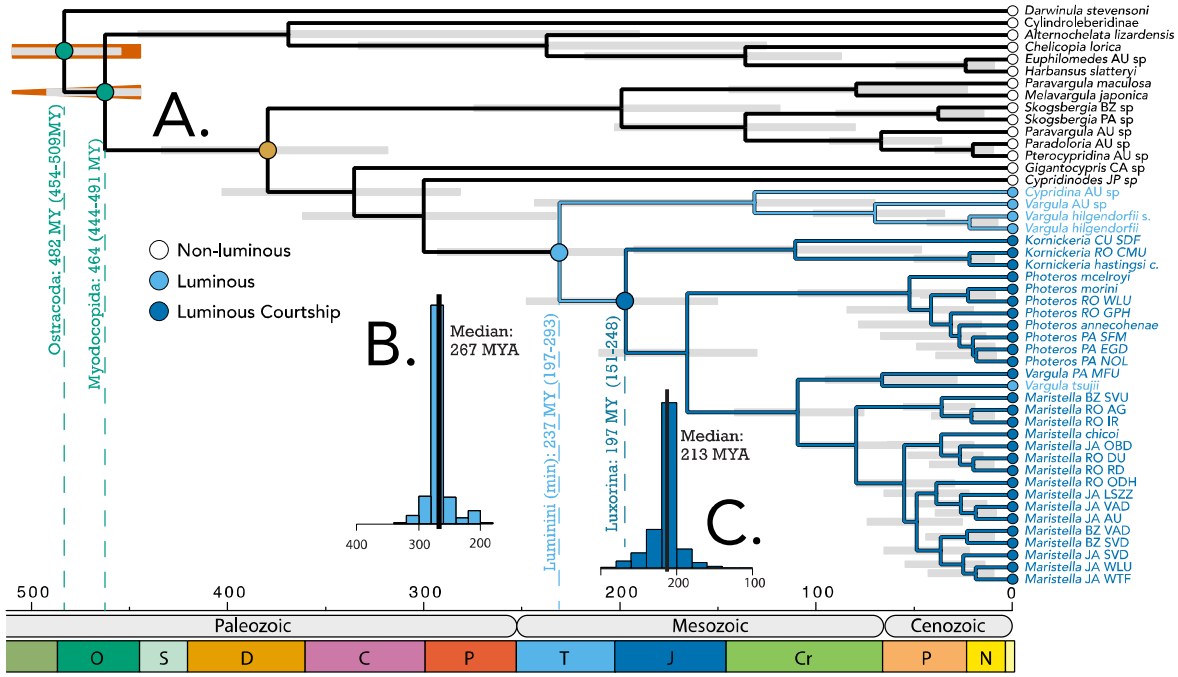


Figure 2. Relaxed molecular clock analysis. Figure 2 copied for more convenient comparison with Fig. S8. We used a Bayesian relaxed molecular clock approach, implemented in MrBayes (Ronquist and Huelsenbeck 2003) with 15 orthologs selected using sortadate (Smith, Brown, and Walker 2018). We used two fossil constraints: Myodocopida, with offset exponential prior with a minimum of 448.8 MY and maximum of 509 MY and Ostracoda with a uniform prior between 443.8 and 509 MY (Oakley et al. 2013; Wolfe et al. 2016) We used the Independent Gamma Rate (IGR) model in MrBayes, (Lepage et al. 2007; Chi Zhang 2016) with a prior of exp(10). We assumed a lognormal distribution with a mean of 0.001 and standard deviation of 0.0007 for the rate prior. We assumed fixed-rate amino acid models with aamodelpr=mixed. We ran 500K steps of Markov Chain Monte Carlo (MCMC) in two chains, using a starting phylogeny based on parsimony. B. Results of time-tree stochastic character mapping (ttscm) estimates origin of bioluminescence to be 267 MYA because that is the median age of character state transitions to bioluminescence C. Similarly, ttscm estimates the origin of luminous courtship to be 213 MYA.

We also find strong support (1.0) for the relationships among courtship signaling genera, with *Kornickeria* sister to a clade containing *Photeros*, *Maristella*, Z-Group, and C-group (Fig 3). Next, *Photeros* is the sister clade (0.76) to *Maristella*, Z-Group, and C-group, with Z-group as the sister of *Maristella* plus C-group (1.0). The non-luminous *Skogsbergia* is not monophyletic in the species tree, but this is not well supported and we hypothesize that adding transcriptome data for more species in this genus would recover monophyly. Species-level relationships are largely consistent across analyses, with a few exceptions, including the placement of *Photeros*

jamescasei and *Maristella chicoi,* which differ in the HaMStR analysis, although with low support (Figure S5,6), compared to ASTRAL-Pro analysis (Fig 1). We included *P. jamescasei* because it is a described species even though the quality of the transcriptome was low. The low number of sequences in the transcriptome probably destabilizes phylogenetic placement, especially with HaMSTtR which uses fewer genes than ASTRAL-Pro. Therefore, we prioritize the relationships found with ASTRAL-Pro because these use larger data sets for phylogenetic estimation by including paralogs.

Divergence Time Estimates

Relaxed molecular clock analyses place the origin of the clade of species with bioluminescent courtship between 151-248 MY (Bayesian Highest Posterior Density) with a median value of 197 MYA (Fig. 2C). Using time tree stochastic character mapping (ttscm) (Alexandrou et al. 2013), we estimate the character state transition to luminous courtship occurred 213 MYA. The estimates for the origin of cypridinid bioluminescence are less clear because of issues with taxon sampling and the low quality of the transcriptome of *Jimmorinia*. Namely, the data set used for divergence times does not contain any species of the clade containing *Jimmorinia*, *Vargula norvegica*, or the other luminous Australian species; these taxa form the sister group to the rest of the luminous cypridinids (Fig. 1). Nevertheless, we can estimate the age of the common ancestor of all luminous species with transcriptome data as a minimum estimate for the origin of cypridinid bioluminescence to be 197-293 MY with a median value of 237. Similarly, we can estimate the age of the common ancestor of luminous cypridinids and their closest non-luminous relative as a maximum age to be 233-361 with a median value of 300 MY. Using time tree stochastic character mapping (ttscm) (Alexandrou et al. 2013), we estimate the origin of bioluminescence occurred 267 MYA (Fig. 2B). We also note our estimates of even the most closely related species are always more than 8 MY (Fig S9), even for two populations of V. hilgendorfii. Although populations of *V. hilgendorfii* are quite diverged (Ogoh and Ohmiya 2005) our estimates seem very old at the tips of the tree and we speculate this could be an artifact of extrapolating young nodes from the deepest nodes in the absence of other suitable fossils.

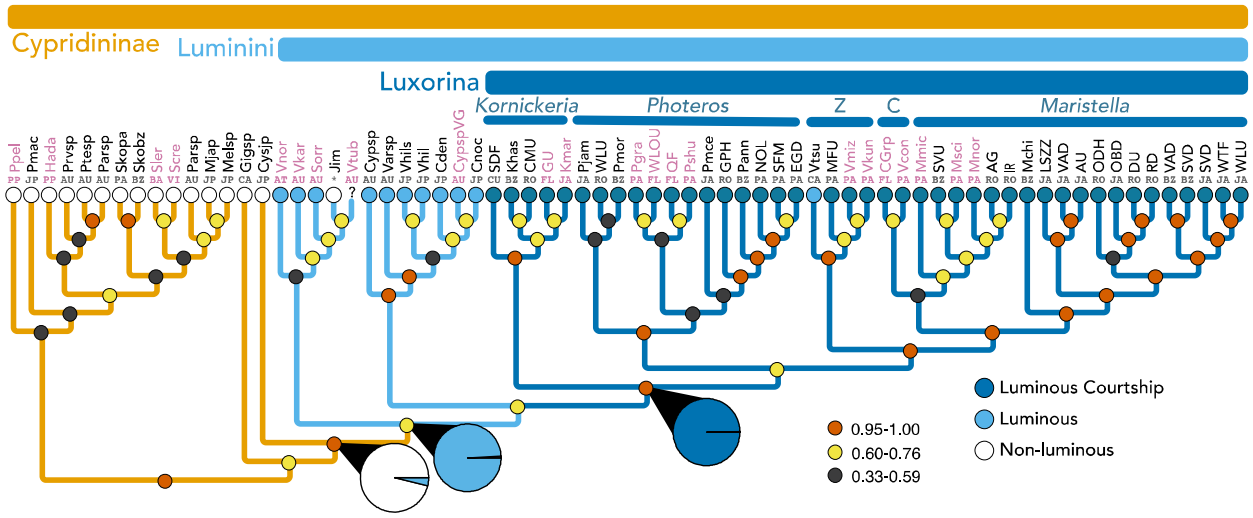


Figure 3. Expanded taxon sampling by adding individual mitochondrial genes, mainly from Torres and Gonzalez (2007) to the transcriptome data in Figure 1. We inferred this species phylogeny using the multi-species-coalescent, employed with ASTRAL-Pro (Chao Zhang et al. 2020). Taxa colored pink are those that contain only mitochondrial data. Locality abbreviations are as in Fig. 1 and also include additional locations of VI=US Virgin Islands and FL=Florida USA. For Jimmorinia (*), we combined data from transcriptomes from a Jamaican species (*J. cf gunnari*) with mitochondrial DNA from a specimen (*Jimmorinia* sp USVI) unidentified to species from VI. Pie charts represent proportions of likelihood values for alternative ancestral states. Bioluminescence was most likely absent (white with hash marks) prior to Luminini, as supported by a significantly high proportion of likelihood for absence of bioluminescence compared to presence (light blue). Significantly high likelihood proportions first appear in Luminini for bioluminescence (light blue) and Luxorina for bioluminescent courtship signaling (dark blue with hashes), compared to absence of those traits (black), which were analyzed separately as binary traits (see Supplement for full ancestral state reconstruction results).

Discussion

The origins of bioluminescence dramatically impacted biodiversity (Ellis and Oakley 2016), and fully understanding these impacts requires a solid phylogenetic foundation. Phylogeny tests homology, demonstrating which species share a common origin of bioluminescence, related structures, and related functions. Establishing sister-group relationships for luminous and non-luminous clades and signaling and non-signaling clades allows for comparisons of species richness associated with the origins of light production or bioluminescent courtship (Davis et al. 2014; Ellis and Oakley 2016). Finally, estimates of divergence times coupled with ancestral state reconstruction provide estimates of when bioluminescence originated, allowing inferences of the biotic and abiotic factors present at the time of transitions in character states. The present study represents the first phylotranscriptomic study of the ostracod family Cypridinidae, with particular focus on species that use bioluminescence for courtship. This new analysis provides inferences about homology of bioluminescence and luminous courtship, and the timing of

character state transitions of cypridinid bioluminescence, luminous courtship, and diversification events.

Homology of Cypridinid Bioluminescence and Luminous Courtship

Although future studies would benefit from high quality genome or transcriptome data from some taxa that are currently unavailable, our current analyses reinforce previous hypotheses about homology of two bioluminescence traits (Cohen and Morin 2003; Torres and Gonzalez 2007; Morin 2019). First, our results support a single origin of cypridinid bioluminescence (Fig 3). For this clade of bioluminescent cypridindids, we propose the name Luminini (lumen = light). Although monophyly of Luminini is well supported in the transcriptomic phylogeny (Fig. 1), we were unable to obtain high quality transcriptomic data from some key species, for which we have only mitochondrial data (Fig. 3). These species may form a clade within Luminini that also includes non-luminous *Jimmorinia*, leading to an inference of loss of bioluminescence in that genus (Figs 3, S14). Also in this clade are luminous species *Vargula norvegica*, *Sheina orri*, and *Vargula karamu*. While the relationships among these species are not well supported based on mitochondrial data alone, there is still reasonable support for Luminini as a whole (Fig. 3). Furthermore, transcriptomic data for *Jimmorinia* strongly support that it is a member of Luminini (Fig 1). Although future work could obtain more data from these species to further test the Luminini hypothesis, there currently exists strong support for this clade based on current and previous work (Cohen and Morin 2003).

For the strongly supported clade of cypridinids that use light for courtship (Figs 1, 3), we propose the sub-tribe name Luxorina (lux= light, uxoriae ~= courtship). Although an important consideration for defining Luxorina, our molecular data unfortunately does not contain the monotypic genus *Enewton* (Cohen and Morin 2010). Despite returning to the same collection site as Cohen and Morin (2010) in Jamaica, we could not find the courtship signaling species *E. harveyi*. In previous parsimony analyses of morphological traits (Cohen and Morin 2003, 2010), *E. harveyi* is the sister of the rest of Luxorina. In the absence of molecular data, *Enewton* as sister to the rest of Luxorina remains the best hypothesis and we propose to include *Enewton* within Luxorina. *Enewton harveyi* occurs in Jamaica and possibly Puerto Rico and Bahamas (Morin and Cohen 2010) and could be targeted for new collections. Alternatively, museum samples could provide genetic data for future studies to test this hypothesis.

Loss of luminous courtship in “Vargula tsujii” *and polyphyletic* Vargula

Vargula tsujii, with no known bioluminescent courtship, is a California species now separated by the Isthmus of Panama from its closest relatives in the Caribbean, which all produce luminous courtship signals. Although clearly nested within Luxorina in previous studies (Cohen and Morin 2003; Gonzalez and Torres 2010), the specific placement of *V. tsujii* varied. Cohen and Morin (2003) inferred *V. tsujii* to be sister to a clade containing *Kornickeria* and *Maristella*, and a variety of other undescribed lineages, whereas Torres and Gonzalez (2007) inferred its position from mitochondrial data as sister to a clade containing Z-group and U-group species. A similar result from our transcriptome analyses is the inference of a sister-group relationship between *Vargula tsujii* and an undescribed signaling species from Panama (MFU) that signals in very shallow waters near mangroves. Our analyses that add taxa with only mitochondrial data also include two Panamanian species in a clade with *V. tsujii* and MFU: "*Vargula*" *kuna* and "*Vargula*" *mizonoma*, which were previously called members of Z-group (Cohen and Morin 2003). Therefore, we report strong support for a clade that likely represents an undescribed genus that includes the Z-group and *Vargula tsujii*. Even though *V. tsujii* is a described species, the genus containing *V. tsujii* and Z-Group should not be named *Vargula* (Cohen and Morin 2003) because the type species for the genus is the distantly related *V. norvegica* (Fig 3). These results reinforce the highly polyphyletic nature of *Vargula* (Cohen and Morin 2003), which led to renaming of species formerly in *Vargula*, now placed in the genera *Kornickeria* (Cohen and Morin 1989), *Enewton (Cohen and Morin 2010)*, *Photeros (Cohen and Morin 2010)*, and *Maristella* (Reda et al. 2019). In addition to re-naming *V. tsujii*, "*V.*" *contragula, "V." kuna, "V." mizonomma and related species called C-group or Z-group* should be renamed because they are also clearly within Luxorina and distantly related (Fig. 3) to the type species *Vargula norvegica*. However, current data from C-group and some members of Z-group is limited to mitochondrial data, which do not confidently resolve their phylogenetic position (Fig 3). Consequently, C-group could be another new genus or included with *Maristella* (Fig 3).

Timing of origins for bioluminescence traits in Cypridinidae

Although previous publications (Cohen and Morin 2003; Morin 2019) hypothesized a recent origin of Luxorina because they only occur in the recently originated (~3.8 MY) Caribbean Sea, our results provide the first explicit divergence time estimates within Cypridinidae. We estimate the origin of Luxorina to be 197 (151- 248) MYA (Fig 2) and the origin of bioluminescent courtship to be 213 MYA (Fig 2), significantly older than the origin of the modern Caribbean created by the closure of the Isthmus of Panama, and broadly consistent with the timing of initial crust formation between the Americas 154-190 MYA (Baumgartner et al. 2013). These results

suggest that oceanographic conditions of the modern Caribbean, such as its warm temperatures, do not explain the restricted distribution of Luxorina and that the origin of courtship signaling corresponded at least roughly in time and place with the splitting of the Americas to form the proto-Caribbean. Mate signaling is likely to have evolved from co-option of the anti-predator display within sexual contexts, as hypothesized by sensory bias and other models, but which needs further testing. In all cases our data provides a strong phylogenetic framework for testing alternative models on the origin of bioluminescent courtship in Luxorina.

Our estimates for the origin of Luminini (more than 197 MYA) and cypridinid bioluminescence (267 MYA) makes it one of the oldest origins of bioluminescence quantified to date. Other ancient origins of bioluminescence may include sepiolid cephalopods at 255 MYA (Ellis and Oakley 2016). However, the homology of bioluminescence is not well resolved in sepiolids (Lindgren et al. 2012; Pankey et al. 2014) and if there are multiple non-homologous origins, their ages could be much younger. Furthermore, more recently published divergence time estimates of sepiolids are younger than 255 MY (Tanner et al. 2017). One reliably old origin of bioluminescence is that of ctenophores. Bioluminescence originated at least 350 ± 88 MYA based on relaxed clock methods, because the trait is probably ancestral and homologous in living ctenophores (Francis et al. 2015) and stem ctenophores are even older (Parry et al. 2021), although whether they were bioluminescent is not clear.

Importantly, difficulties in taxon sampling complicate divergence time estimates for bioluminescence traits of cypridinids. If *Enewton* is the sister group of all other Luxorina, this pushes back the origin of luminous courtship. Similarly, including species like *Vargula norvegica* and *Sheina orri* in divergence estimates adds an older node within Luminini. However, because the ttscm technique does not estimate character state changes at nodes only, it may be less affected by taxon sampling. But even the ttscm estimates of 213 MYA for courtship and 267 MYA for bioluminescence could be older with more complete sampling. Therefore, the most conservative interpretation of our results is to use the 95% confidence intervals of critical nodes to make minimum estimates for character state transitions. Assuming such minimum estimates, the origin of courtship signaling is at least 151 MYA and the origin of bioluminescence is at least 197 MYA in cypridinids. Even using these very conservative minima, the origin of courtship signaling is much older than the closing of the Isthmus of Panama and Luminini is one of the oldest origins of bioluminescence quantified thus far.

Genera of Luxorina

The relationships among genera in Luxorina varied among previous studies. One hypothesis suggested a close relationship between *Kornickeria* and *Maristella,* supported by reproductive (slow-flashing courtship signals) and morphological characters (Cohen and Morin 1993; Morin and Cohen 2010, 2017; Reda et al. 2019). Previous phylogenetic analyses based on mitochondrial data alone were consistent with this hypothesis, albeit with low support (Torres and Gonzalez 2007; Hensley et al. 2019). In contrast, our results do not support a sister relationship between *Maristella* and *Kornickeria*, instead supporting *Kornickeria* as the sister-group to the rest of Luxorina*.* Interestingly, our topology based on transcriptomes is consistent with the phylogeny of c-luciferase, the light-catalyzing protein (Hensley et al. 2021), which is not in our final ortholog set and therefore independent of the analyses presented here. Based on these considerations, we favor *Kornickeria* as the sister-group to the rest of Luxorina*.*

Future Directions and Summary

Although our analyses contain a broad sampling across Cypridinidae, future studies should include representatives of all genera, including *Azygocypridina, Codonocera, Doloria, Enewton, Isocypridina, Lowrya, Metavargula, Rheina, Rugosidoloria,* and *Siphonostra.* Several are monotypic and difficult to collect live, such as *Rheina*, which inhabits abyssal depths (Kornicker 1989). In addition to genus-level representatives, some key taxa of Luminini have only mitochondrial data or low quality transcriptomic data, including *Jimmorinia*, "*Vargula*" *contragula* and C-group, "*V.*" *norvegica*, "*V*". *tubulata*, and "*V.*" *karamu*. Some sequencing methods such as Anchored Hybrid Enrichment (AHE) (Lemmon et al. 2012) allow use of museum collections and could prove useful to add molecular data for cypridinid species that are difficult to collect live.

Bioluminescence evolved in the family Cypridinidae independently from other animals, and cypridinids use a substrate, cypridinid-luciferin (Morin 2011), that is endogenous and chemically different from that of other bioluminescent organisms (Hastings 1983; Thompson et al. 1989; Kato et al. 2004, 2007). Therefore, Luminini represent a critical lineage for understanding how bioluminescence evolved (Haddock et al. 2010; Morin 2019; Goodheart et al. 2020; Lau and Oakley 2020). The phylogenetic framework including divergence time estimates provides a foundation for important questions related to molecular evolution and features important for rapid diversification, including those related to behavior (Boake et al. 2002), sexual selection (Ritchie 2007), species diversification (Ellis and Oakley 2016), and the influence of the biochemical properties of bioluminescence on evolution (Hensley et al. 2019).

Acknowledgments

We acknowledge funding support from the National Science Foundation to THO (DEB-1457754, DEB-1146337), ET (DEB-1457462), GAG (DEB-1457439) and EAE (1515576 and 1702011). Permits from the Jamaican National Environment and Planning Agency (Permit Ref. no. 18/27), the Belize Fisheries Department (Permit no. 000003-16), the Honduran Department of Fish and Wildlife (Permit no. DE-MO-082-2016), the Puerto Rican Department of Natural and Environmental Resources (DRNA; Permit no. 2016-IC-113), and Panamanian Ministry of the Environment (MiAMBIENTE; Permit no. SE/A-33-17) were obtained for all animal collections from their respective locations. We thank Drs. Karen Osborn and Jamie Fergus for sending us the *Gigantocypris* sample. We also thank Drs. Yasuo Mitani, Yoshihiro Ohmiya, and Katsunori Ogoh for their assistance in collecting Japanese ostracods. Thanks to Andrew Parker for sending samples from Australia and Anne Cohen for much wisdom shared through countless discussions with all of us and for confirming identification of some species.

References

Alexandrou M.A., Swartz B.A., Matzke N.J., Oakley T.H. 2013. Genome duplication and multiple evolutionary origins of complex migratory behavior in Salmonidae. Mol. Phylogenet. Evol. 69:514–523.

Bacon C.D., Silvestro D., Jaramillo C., Smith B.T., Chakrabarty P., Antonelli A. 2015. Biological evidence supports an early and complex emergence of the Isthmus of Panama. Proc. Natl. Acad. Sci. U. S. A. 112:6110–6115.

Bailey W.J. 2003. Insect duets: underlying mechanisms and their evolution. Physiol. Entomol. 28:157–174.

Baumgartner P.O., Rojas-Agramonte Y., Sandoval-Gutierrez M., Urbani F., García-Delgado D., Garban G., Pérez Rodríguez M. 2013. Late Jurassic breakup of the Proto-Caribbean and circum-global circulation across Pangea. :13408.

Beaulieu J.M., O'Meara B.C., Donoghue M.J. 2013. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. Syst. Biol. 62:725–737.

Bernt M., Donath A., Jühling F., Externbrink F., Florentz C., Fritzsch G., Pütz J., Middendorf M., Stadler P.F. 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. 69:313–319.

Boake C.R.B., Arnold, Breden, Meffert, Ritchie, Taylor, Wolf, Moore. 2002. Genetic Tools for Studying Adaptation and the Evolution of Behavior. The American Naturalist. 160:S143.

Branchini B.R., Southworth T.L., Murtiashaw M.H., Magyar R.A., Gonzalez S.A., Ruggiero M.C., Stroh J.G. 2004. An alternative mechanism of bioluminescence color determination in firefly luciferase. Biochemistry. 43:7255–7262.

Capella-Gutierrez S., Silla-Martinez J.M., Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 25:1972–1973.

Coates A.G., Jackson J.B.C., Collins L.S., Cronin T.M., Dowsett H.J., Bybell L.M., Jung P., Obando J.A. 1992. Closure of the Isthmus of Panama: The near-shore marine record of Costa Rica and western Panama. Geological Society of America Bulletin. 104:814–828.

Cohen A.C., Morin J.G. 1986. Three new luminescent ostracodes of the genus Vargula (Myodocopida, Cypridinidae) from the San Blas region of Panama. Natural History Museum of Los Angeles County.

Cohen A.C., Morin J.G. 1989. Six New Luminescent Ostracodes of the Genus Vargula (Myodocopida: Cypridinidae) From the San Blas Region of Panama. J. Crustacean Biol. 9:297–340.

Cohen A.C., Morin J.G. 1990a. Patterns of Reproduction in Ostracodes: A Review. J. Crustacean Biol. 10:184–212.

Cohen A.C., Morin J.G. 1990b. Morphological relationships of bioluminescent Caribbean species of Vargula (Myodocopa). Ostracoda and Global Events.:381–400.

Cohen A.C., Morin J.G. 1993. The cypridinid copulatory limb and a new genus Kornickeria (Ostracoda: Myodocopida) with four new species of bioluminescent ostracods from the Caribbean. Zool. J. Linn. Soc. 108:23–84.

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

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

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

Davis M.P., Holcroft N.I., Wiley E.O., Sparks J.S., Leo Smith W. 2014. Species-specific bioluminescence facilitates speciation in the deep sea. Mar. Biol. 161:1139–1148.

Ebersberger I., Strauss S., von Haeseler A. 2009. HaMStR: profile hidden markov model based search for orthologs in ESTs. BMC Evol. Biol. 9:157.

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

Emms D.M., Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20:238.

Farris D.W., Jaramillo C., Bayona G., Restrepo-Moreno S.A., Montes C., Cardona A., Mora A., Speakman R.J., Glascock M.D., Valencia V. 2011. Fracturing of the Panamanian Isthmus during initial collision with South America. Geology. 39:1007–1010.

Francis W.R., Shaner N.C., Christianson L.M., Powers M.L., Haddock S.H.D. 2015. Occurrence of Isopenicillin-N-Synthase Homologs in Bioluminescent Ctenophores and Implications for Coelenterazine Biosynthesis. PLoS One. 10:e0128742.

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

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

Gonzalez V.L., Torres E. 2010. Phylogenetic analysis of the ostracod family Cypridinidae and the evolution of bioluminescence. CLADISTICS. 26:212–212.

Goodheart J.A., Minsky G., Brynjegard-Bialik M.N., Drummond M.S., Munoz J.D., Fallon T.R., Schultz D.T., Weng J.-K., Torres E., Oakley T.H. 2020. Laboratory culture of the California Sea Firefly Vargula tsujii (Ostracoda: Cypridinidae): Developing a model system for the evolution of marine bioluminescence. Sci. Rep. 10:10443.

Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N., Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29:644–652.

Haas B.J., Papanicolaou A., Yassour M., Grabherr M., Blood P.D., Bowden J., Couger M.B., Eccles D., Li B., Lieber M., MacManes M.D., Ott M., Orvis J., Pochet N., Strozzi F., Weeks N., Westerman R., William T., Dewey C.N., Henschel R., LeDuc R.D., 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 S.H.D., Moline M.A., Case J.F. 2010. Bioluminescence in the sea. Ann. Rev. Mar. Sci. 2:443–493.

Hastings J.W. 1983. Biological diversity, chemical mechanisms, and the evolutionary origins of bioluminescent systems. J. Mol. Evol. 19:309–321.

Hensley N.M., Ellis E.A., Gerrish G.A., Torres E., Frawley J.P., Oakley T.H., Rivers T.J. 2019. Phenotypic evolution shaped by current enzyme function in the bioluminescent courtship signals of sea fireflies. Proc. Biol. Sci. 286:20182621.

Hensley N.M., Ellis E.A., Leung N.Y., Coupart J., Mikhailovsky A., Taketa D.A., Tessler M., Gruber D.F., De Tomaso A.W., Mitani Y., Rivers T.J., Gerrish G.A., Torres E., Oakley T.H. 2021. Selection, drift, and constraint in cypridinid luciferases and the diversification of bioluminescent signals in sea fireflies. Mol. Ecol. 30:1864–1879.

Huelsenbeck J.P., Nielsen R., Bollback J.P. 2003. Stochastic mapping of morphological characters. Syst. Biol. 52:131–158.

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

Kato S.-I., Oba Y., Ojika M. 2007. Biosynthesis of cypridina luciferin in Cypridina noctiluca. Heterocycles. 72:673–676.

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

11434.

Kocot K.M., Citarella M.R., Moroz L.L., Halanych K.M. 2013. PhyloTreePruner: A Phylogenetic Tree-Based Approach for Selection of Orthologous Sequences for Phylogenomics. Evol. Bioinform. Online. 9:429–435.

Kornicker L.S. 1989. Bathyal and Abyssal Myodocopid Ostracoda of the Bay of Biscay and Vicinity. Smithsonian Contributions to Zoology.:1–134.

Krueger F. 2012. Trim Galore: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type (Reduced Representation Bisufite-Seq) libraries. URL http://www. bioinformatics. babraham. ac. uk/projects/trim_galore/. (Date of access: 28/04/2016).

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

Lanfear R., Calcott B., Kainer D., Mayer C., Stamatakis A. 2014. Selecting optimal partitioning schemes for phylogenomic datasets. BMC Evol. Biol. 14:82.

Lanfear R., Frandsen P.B., Wright A.M., Senfeld T., Calcott B. 2016. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. Molecular Biology and Evolution.:msw260.

Lau E.S., Oakley T.H. 2020. Multi-level convergence of complex traits and the evolution of bioluminescence. .

Lemmon A.R., Emme S.A., Lemmon E.M. 2012. Anchored Hybrid Enrichment for Massively High-Throughput Phylogenomics. Syst. Biol. 61:727–744.

Lepage T., Bryant D., Philippe H., Lartillot N. 2007. A general comparison of relaxed molecular clock models. Mol. Biol. Evol. 24:2669–2680.

Lindgren A.R., Pankey M.S., Hochberg F.G., Oakley T.H. 2012. A multi-gene phylogeny of Cephalopoda supports convergent morphological evolution in association with multiple habitat shifts in the marine environment. BMC Evol. Biol. 12.

Minh B.Q., Hahn M.W., Lanfear R. 2020. New Methods to Calculate Concordance Factors for Phylogenomic Datasets. Mol. Biol. Evol. 37:2727–2733.

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

Morin J.G. 1986. Firefleas of the Sea: Luminescent Signaling in Marine Ostracode Crustaceans. The Florida Entomologist. 69:105.

Morin J.G. 2011. Based on a review of the data, use of the term “cypridinid” solves the Cypridina/Vargula dilemma for naming the constituents of the luminescent system of ostracods in the family Cypridinidae. Luminescence. 26:1–4.

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

Morin J.G., Cohen A.C. 1988. Two new luminescent ostracodes of the genus Vargula (Myodocopida, Cypridinidae) from the San Blas region of Panama. J. Crustacean Biol. 8:620–638.

Morin J.G., Cohen A.C. 2010. It’s All About Sex: Bioluminescent Courtship Displays, Morphological Variation and Sexual Selection in Two New Genera of Caribbean Ostracodes. Journal of Crustacean Biology. 30:56–67.

Morin J.G., Cohen A.C. 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.

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

Oakley T.H., Alexandrou M.A., Ngo R., Pankey M.S., Churchill C.K.C., Chen W., Lopker K.B. 2014. Osiris: accessible and reproducible phylogenetic and phylogenomic analyses within the Galaxy workflow management system. BMC Bioinformatics. 15:230.

Oakley T.H., Wolfe J.M., Lindgren A.R., Zaharoff A.K. 2013. Phylotranscriptomics to bring the understudied into the fold: monophyletic ostracoda, fossil placement, and pancrustacean phylogeny. Mol. Biol. Evol. 30:215–233.

O’Dea A., Lessios H.A., Coates A.G., Eytan R.I., Restrepo-Moreno S.A., Cione A.L., Collins L.S., de Queiroz A., Farris D.W., Norris R.D., Stallard R.F., Woodburne M.O., Aguilera O.,

Aubry M.-P., Berggren W.A., Budd A.F., Cozzuol M.A., Coppard S.E., Duque-Caro H., Finnegan S., Gasparini G.M., Grossman E.L., Johnson K.G., Keigwin L.D., Knowlton N., Leigh E.G., Leonard-Pingel J.S., Marko P.B., Pyenson N.D., Rachello-Dolmen P.G., Soibelzon E., Soibelzon L., Todd J.A., Vermeij G.J., Jackson J.B.C. 2016. Formation of the Isthmus of Panama. Sci Adv. 2:e1600883.

Ogoh K., Ohmiya Y. 2004. Complete mitochondrial DNA sequence of the sea-firefly, Vargula hilgendorfii (Crustacea, Ostracoda) with duplicate control regions. Gene. 327:131–139.

Ogoh K., Ohmiya Y. 2005. Biogeography of luminous marine ostracod driven irreversibly by the Japan current. Mol. Biol. Evol. 22:1543–1545.

Pagel M. 1999. The Maximum Likelihood Approach to Reconstructing Ancestral Character States of Discrete Characters on Phylogenies. Systematic Biology. 48:612–622.

Pankey M.S., Minin V.N., Imholte G.C., Suchard M.A., Oakley T.H. 2014. Predictable transcriptome evolution in the convergent and complex bioluminescent organs of squid. Proceedings of the National Academy of Sciences. 111:E4736–E4742.

Paradis E., Claude J., Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 20:289–290.

Parry L.A., Lerosey-Aubril R., Weaver J.C., Ortega-Hernández J. 2021. Cambrian comb jellies from Utah illuminate the early evolution of nervous and sensory systems in ctenophores. iScience.:102943.

Pham H.T.M., Tanaka H., Karanovic I. 2020. Molecular and Morphological Diversity of Heterodesmus Brady and Its Phylogenetic Position within Cypridinidae (Ostracoda). Zoological Science. 37:240.

Price M.N., Dehal P.S., Arkin A.P. 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One. 5:e9490.

Reda N.J., Morin J.G., Torres E., Cohen A.C., Schawaroch V., Gerrish G.A. 2019. Maristella, a new bioluminescent ostracod genus in the Myodocopida (Cypridinidae). Zool. J. Linn. Soc.

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

Ritchie M.G. 2007. Sexual Selection and Speciation. Annu. Rev. Ecol. Evol. Syst. 38:79–102.

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

Rivers T.J., Morin J.G. 2012. The relative cost of using luminescence for sex and defense: light budgets in cypridinid ostracods. J. Exp. Biol. 215:2860–2868.

Rivers T.J., Morin J.G. 2013. Female ostracods respond to and intercept artificial conspecific male luminescent courtship displays. Behavioral Ecology. 24:877–887.

Ronquist F., Huelsenbeck J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.

Schön I., Rodriguez F., Dunn M., Martens K., Shribak M., Arkhipova I.R. 2021. A Survey of Transposon Landscapes in the Putative Ancient Asexual Ostracod Darwinula stevensoni. Genes . 12.

Schwentner M., Richter S., Rogers D.C., Giribet G. 2018. Tetraconatan phylogeny with special focus on Malacostraca and Branchiopoda: highlighting the strength of taxon-specific matrices in phylogenomics. Proc. Biol. Sci. 285.

Shimomura O., Johnson F.H. 1971. Mechanism of the luminescent oxidation of cypridina luciferin. Biochem. Biophys. Res. Commun. 44:340–346.

Simão F.A., Waterhouse R.M., Ioannidis P., Kriventseva E.V., Zdobnov E.M. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 31:3210–3212.

Simion P., Belkhir K., François C., Veyssier J., Rink J.C., Manuel M., Philippe H., Telford M.J. 2018. A software tool “CroCo” detects pervasive cross-species contamination in next generation sequencing data. BMC Biol. 16:28.

Smith S.A., Brown J.W., Walker J.F. 2018. So many genes, so little time: A practical approach to divergence-time estimation in the genomic era. PLoS One. 13:e0197433.

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

Tanner A.R., Fuchs D., Winkelmann I.E., Gilbert M.T.P., Pankey M.S., Ribeiro Â.M., Kocot K.M., Halanych K.M., Oakley T.H., da Fonseca R.R., Pisani D., Vinther J. 2017. Molecular clocks indicate turnover and diversification of modern coleoid cephalopods during the Mesozoic Marine Revolution. Proc. Biol. Sci. 284.

Thalén F. 2018. PhyloPyPruner: Tree-based Orthology Inference for Phylogenomics with New Methods for Identifying and Excluding Contamination. .

Thompson E.M., Nagata S., Tsuji F.I. 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.

Tinn O., Oakley T.H. 2008. Erratic rates of molecular evolution and incongruence of fossil and molecular divergence time estimates in Ostracoda (Crustacea). Mol. Phylogenet. Evol. 48:157–167.

Torres E., Gonzalez V.L. 2007. Molecular Phylogeny of Cypridinid Ostracodes and the Evolution of Bioluminescence. Proceedings of the 14th International Symposium on Bioluminescence and Chemiluminescence: Chemistry, Biology, and Applications.:269–272.

Wakayama N., Abe K. 2006. The evolutionary pathway of light emission in myodocopid Ostracoda. Biol. J. Linn. Soc. Lond. 87:449–455.

Wang X., Xu Q., Xiao J., Miao X., Liu P., Wang Z. 2019. First record of the complete mitochondrial genome of Cypridina dentata (Myodocopida: Cypridinidae). Mitochondrial DNA Part B. 4:1607–1608.

Widder E.A. 2010. Bioluminescence in the ocean: origins of biological, chemical, and ecological diversity. Science. 328:704–708.

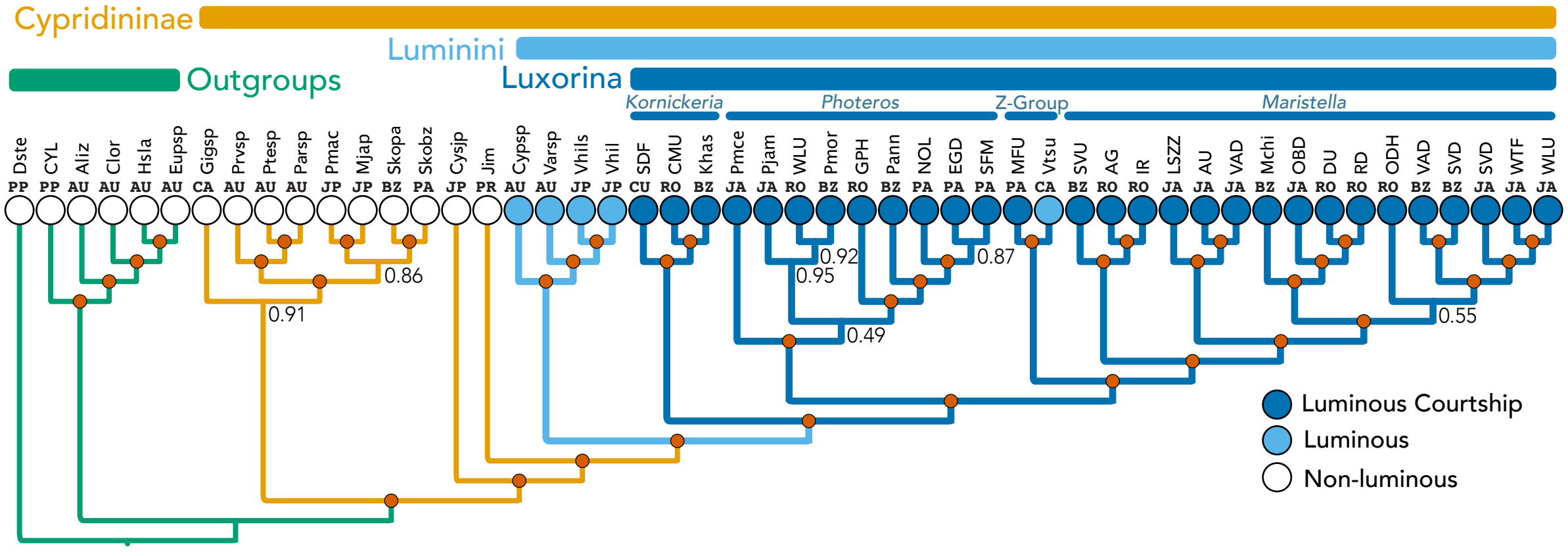
Wilson T., Hastings J.W. 1998. BIOLUMINESCENCE. Annual Review of Cell and Developmental Biology. 14:197–230.

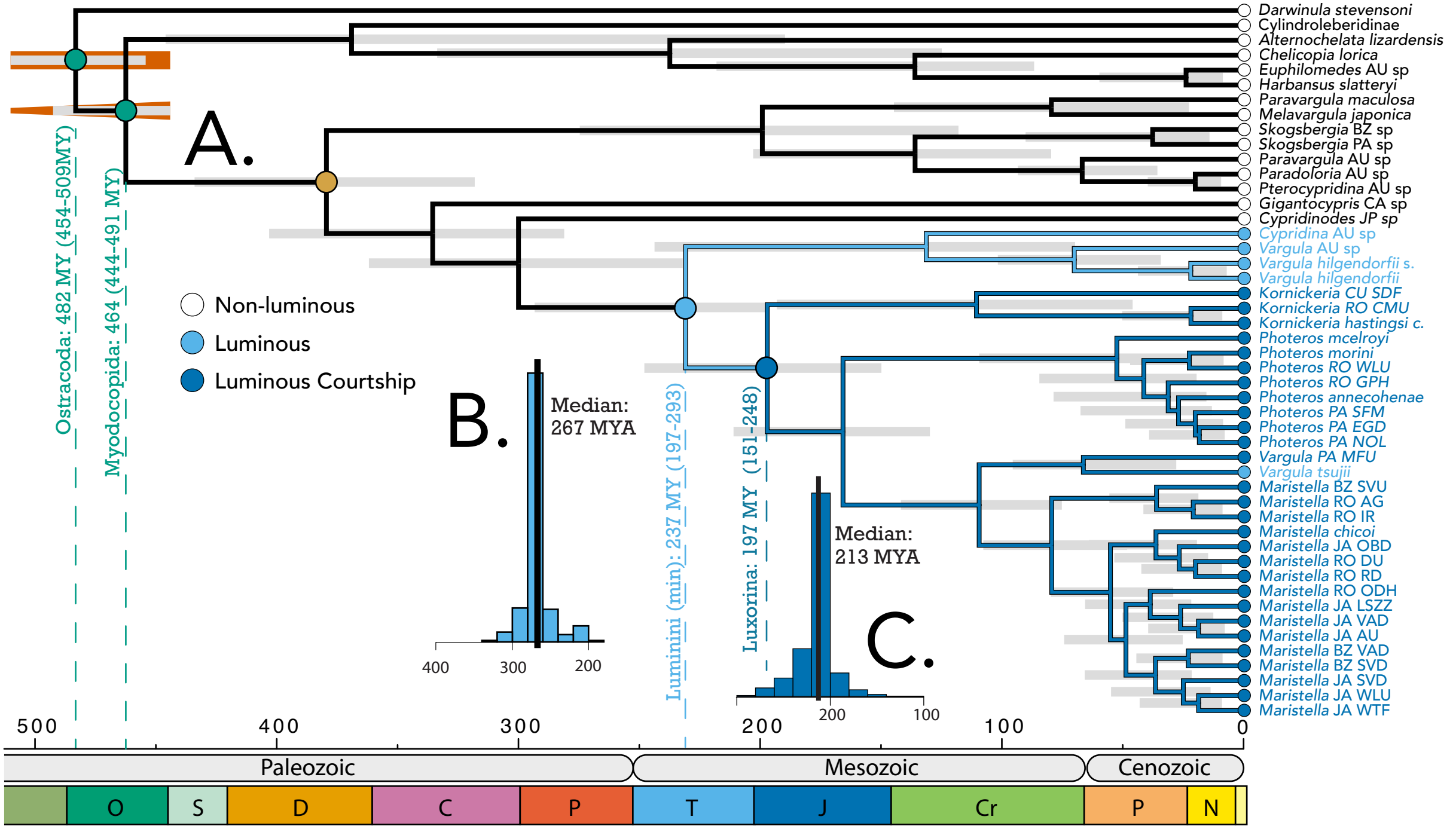
Wolfe J.M., Daley A.C., Legg D.A., Edgecombe G.D. 2016. Fossil calibrations for the arthropod Tree of Life. Earth-Sci. Rev.

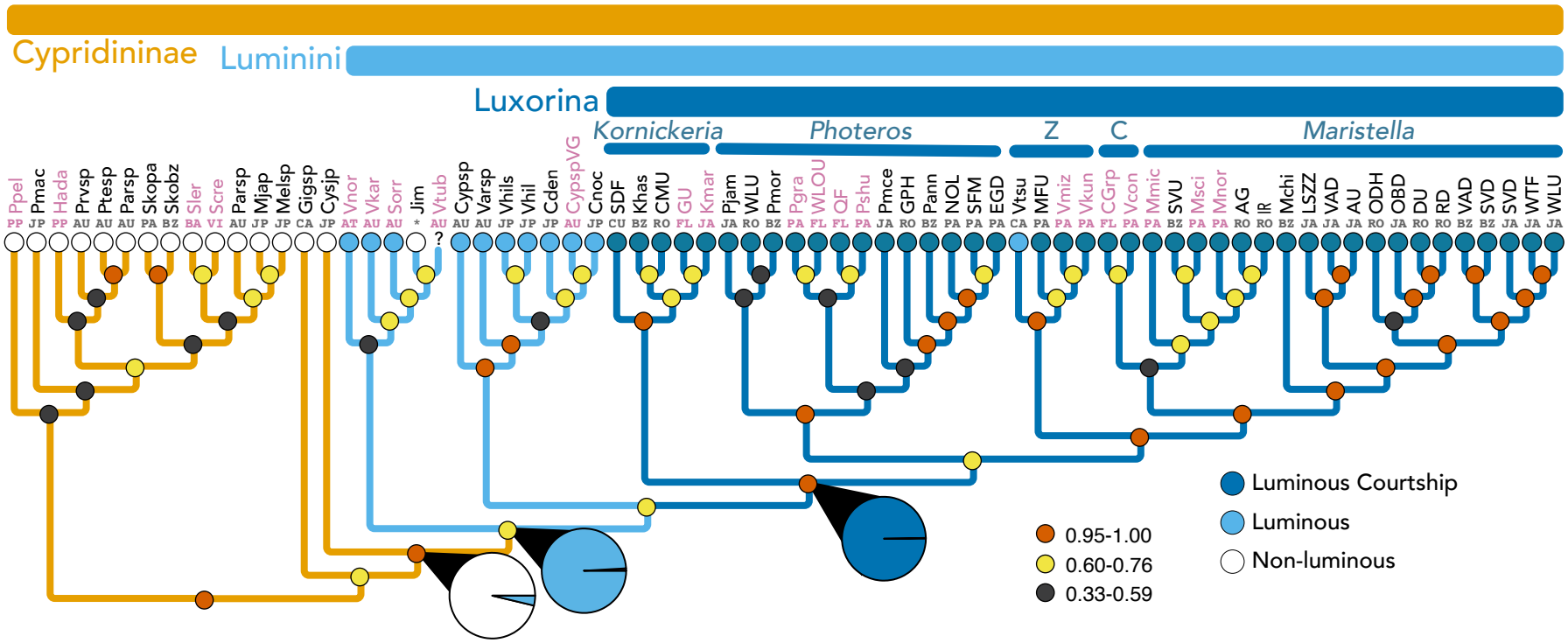
Zhang C. 2016. Molecular Clock Dating using MrBayes. arXiv [q-bio.PE].

Zhang C., Scornavacca C., Molloy E.K., Mirarab S. 2020. ASTRAL-Pro: Quartet-Based

Species-Tree Inference despite Paralogy. Mol. Biol. Evol. 37:3292–3307.


Cypridininae
Luminini
Outgroups
Luxorina
Kornickeria
Photeros
Z-Group
Maristella
Dste PP
CYL PP
Aliz AU
Clor AU
Hsla AU
Eupsp AU
Gigsp CA
Prvsp AU
Ptesp AU
Parsp AU
Pmac JP
Mjap JP
Skopa BZ
Skobz PA
Cysjp JP
Jim PR
Cypsp AU
Varsp AU
Vhils JP
Vhil JP
SDF CU
CMU RO
Khas BZ
Pmce JA
Pjam JA
WLU RO
Pmor BZ
GPH RO
Pann BZ
NOL PA
EGD PA
SFM PA
MFU PA
Vtsu CA
SVU BZ
AG RO
IR RO
LSZZ JA
AU JA
VAD JA
Mchi BZ
OBD JA
DU RO
RD RO
ODH RO
VAD BZ
SVD BZ
SVD JA
WTF JA
WLU JA
0.86
0.91
0.92
0.95
0.87
0.49
0.55
Luminous Courtship
Luminous
Non-luminous


A.
B.
C.
Ostracoda: 482 MY (454-509MY)
Myodocopida: 464 (444-491 MY)
Luminini (min): 237 MY (197-293)
Luxorina: 197 MY (151-248)
Non-luminous
Luminous
Luminous Courtship
Median: 267 MYA
Median: 213 MYA
400
300
200
200
100
500
400
300
200
100
0
Paleozoic
Mesozoic
Cenozoic
O
S
D
C
P
T
J
Cr
P
N
Darwinula stevensoni
Cylindroleberidinae
Alternochelata lizardensis
Chelicopia lorica
Euphilomedes AU sp
Harbansus slatteryi
Paravargula maculosa
Melavargula japonica
Skogsbergia BZ sp
Skogsbergia PA sp
Paravargula AU sp
Paradoloria AU sp
Pterocypridina AU sp
Gigantocypris CA sp
Cypridinodes JP sp
Cypridina AU sp
Vargula AU sp
Vargula hilgendorfii s.
Vargula hilgendorfii
Kornickeria CU SDF
Kornickeria RO CMU
Kornickeria hastingsi c.
Photeros mcelroyi
Photeros morini
Photeros RO WLU
Photeros RO GPH
Photeros annecohenae
Photeros PA SFM
Photeros PA EGD
Photeros PA NOL
Vargula PA MFU
Vargula tsujii
Maristella BZ SVU
Maristella RO AG
Maristella RO IR
Maristella chicoi
Maristella JA OBD
Maristella RO DU
Maristella RO RD
Maristella RO ODH
Maristella JA LSZZ
Maristella JA VAD
Maristella JA AU
Maristella BZ VAD
Maristella BZ SVD
Maristella JA SVD
Maristella JA WLU
Maristella JA WTF


Cypridininae
Luminini
Luxorina
Kornickeria
Photeros
Z
C
Maristella
Ppel PP
Pmac JP
Hada PP
Prvsp AU
Ptesp AU
Parsp AU
Skopa PA
Skobz BZ
Sler BA
Scre VI
Parsp AU
Mjap JP
Melsp JP
Gigsp CA
Cysjp JP
Vnor AT
Vkar AU
Sorr AU
Jim *
Vtub AU
Cypsp AU
Varsp AU
Vhils JP
Vhil JP
Cden JP
CypspVG AU
Cnoc JP
SDF CU
Khas BZ
CMU RO
GU FL
Kmar JA
Pjam JA
WLU RO
Pmor BZ
Pgra PA
WLOU FL
QF FL
Pshu PA
Pmce JA
GPH RO
Pann BZ
NOL PA
SFM PA
EGD PA
Vtsu CA
MFU PA
Vmiz PA
Vkun PA
CGrp FL
Vcon PA
Mmic PA
SVU BZ
Msci PA
Mnor PA
AG RO
IR RO
Mchi BZ
LSZZ JA
VAD JA
AU JA
ODH RO
OBD JA
DU RO
RD RO
VAD BZ
SVD BZ
SVD JA
WTF JA
WLU JA
0.95-1.00
0.60-0.76
0.33-0.59
Luminous Courtship
Luminous
Non-luminous