

SCIENTIFIC REPORTS



OPEN

Transcriptome sequencing and phylogenetic analysis of four species of luminescent beetles

Kai Wang, Wei Hong, Hengwu Jiao & Huabin Zhao

The evolution of bioluminescence has prompted scientific attention to illuminate phylogenetic relationships of luminescent beetles. However, genomic resources are virtually lacking in rhagophthalmids (Rhagophthalmidae) and their related firefly beetles lampyrids (Lampyridae). Here, we employed the Illumina HiSeq 2000 platform and sequenced the whole-body transcriptomes of the four luminescent beetles: one rhagophthalmid (*Rhagophthalmus* sp.) and three fireflies (*Asymmetricata circumdata*, *Aquatica ficta*, and *Pyrocoelia pectoralis*). We obtained 55.4, 43.4, 38.6, and 36.7 million clean reads for the four species, respectively. All reads were assembled into contigs from which unigenes were derived. All unigenes were annotated by publicly available databases, and a total of 4325 orthologous genes were identified. Using multiple phylogenetic approaches, our transcriptome data confirmed the distinctiveness of Rhagophthalmidae from Lampyridae, which was also supported by our mitogenome analysis using three newly determined mitogenome sequences and 12 previously published ones. Together, this study is the first report of whole transcriptome sequencing data in Rhagophthalmidae and Lampyridae species, representing a valuable genomic resource for studying the origin and evolution of some remarkable traits in these beetles such as bioluminescence. Moreover, our transcriptome and mitogenome data provide useful phylogenetic information that could be of importance in future studies of phylogenetic inference.

Bioluminescence is among the most spectacular features in living organisms, including numerous species of marine fishes, marine invertebrates, terrestrial invertebrates, fungi, bacteria, and protists^{1,2}. The uses of bioluminescence in nature involve a range of vital functions: camouflage, attraction, defense, warning, communication, mimicry, and illumination². The chemical basis of the natural light-producing molecules has been elucidated in the last century³, permitting discoveries of countless valuable applications in biology and medicine using luciferase-based systems². In addition, industrial designers have been ambitious in utilizing the natural light-producing systems in bioluminescent organisms for decoration and street lighting⁴.

Among these bioluminescent organisms, luminescent insects were primarily found in members of the three orders: Collembola (springtails), Diptera (flies), and Coleoptera (beetles)⁵. Within the superfamily Elateroidea of the order Coleoptera, several groups of beetles are able to produce and emit light such as fireflies (Lampyridae), railroad worms (Phengodidae), click beetles (Elateridae), and Rhagophthalmidae^{5,6}. The origin and evolution of bioluminescence in these beetles has prompted a number of studies to illuminate the phylogenetic relationships of these bioluminescent beetles^{7–16}. Although both morphological^{10–12, 14, 16} and molecular^{7–9, 13, 15, 17} features were involved in these studies, molecular evidence has been expected to resolve taxonomic status and phylogenetic relationships that have been contentious based on morphological evidence¹⁸, because the evolutionary history of morphological features is usually complex^{18, 19}. Unfortunately, the incongruence of molecular evidence has also been observed in phylogenetic analyses concerning the phylogenetic relationships of these beetles^{7–9, 13, 15}, possibly because a single or a few genes lack sufficient phylogenetic signals²⁰. For instance, Rhagophthalmidae had previously been assigned to the family Phengodidae inferred from morphological data¹¹; other studies based on morphological, embryological and molecular evidence had assigned Rhagophthalmidae to be a subfamily or genus in the family Lampyridae^{8, 14, 15, 21}, but more recent evidence from both morphological and molecular data had supported the familial status of Rhagophthalmidae^{7, 9, 10, 13, 16, 22–24}, which is distinct from both Lampyridae and Phengodidae. In addition, the taxonomic studies focusing on Lampyridae and Rhagophthalmidae have

Department of Ecology and Hubei Key Laboratory of Cell Homeostasis, College of Life Sciences, Wuhan University, Wuhan, 430072, China. Correspondence and requests for materials should be addressed to H.Z. (email: huabinzhao@whu.edu.cn)

Species	<i>Rhagophthalmus</i> sp.	<i>A. circumdata</i>	<i>A. ficta</i>	<i>P. pectoralis</i>
Clean reads				
Number of reads	55,374,194	43,437,504	38,642,102	36,663,932
Number of bases (nt)	5,496,107,276	3,991,132,686	3,909,904,630	3,532,502,346
Mean length (nt)	99	96	96	96
Contigs				
Number	44,883	57,254	71,424	80,017
Mean length (nt)	1,461	1,087	1,166	1,283
N50 statistics (nt)	2,981	2,016	2,461	2,622
Unigenes				
Number	15,418	24,275	31,520	31,356
Mean length (nt)	1,451	958	838	923
N50 statistics (nt)	2,197	1,556	1,437	1,690

Table 1. Statistics of clean reads, assembled contigs and unigenes of the four transcriptomes generated in this study. Abbreviation: nt, nucleotide(s).

been scarce, especially for species distributed in China^{15, 25, 26}. The scarcity calls for additional data of taxonomic importance. Phylogenomics, i.e. phylogenetic analysis involving genome-scale data, has been believed to out-compete single-gene phylogenetics, which frequently yielded conflicting results caused by stochastic errors from small-scale data sets²⁰. However, genomic resources are extremely limited in Rhagophthalmidae and its related beetles, which may prevent their phylogenetic relationships from in-depth investigations.

To provide novel genomic resources from Rhagophthalmidae and its related firefly beetles (Lampyridae), we employed the Illumina HiSeq 2000 platform and sequenced the whole-body transcriptomes of the four beetles: one rhagophthalmid beetle (*Rhagophthalmus* sp.) and three representatives of lampyrid beetles (*Asymmetricata circumdata*, *Aquatica ficta*, and *Pyrocoelia pectoralis*). To validate the phylogenetic inference based on transcriptome data, we employed the traditional Sanger sequencing to obtain two complete mitogenomes (*Aquatica ficta* and *A. wuhana*) and one nearly complete mitogenome (*Lamprigera yunnana*); we also identified the 13 protein coding genes (PCGs) in mitogenomes from each of the four whole-body transcriptomes (Supplementary Table S1); and we additionally retrieved 12 published complete mitogenomes from members of Rhagophthalmidae, Lampyridae, Phengodidae, Elateridae, Lycidae, Cantharidae, and Tenebrionidae (see the species names and GenBank accessions in Materials and Methods). Our transcriptome analysis recognized the distinctiveness of Rhagophthalmidae from Lampyridae, which was also supported by our mitogenome analysis.

Results

Transcriptome sequencing and assembly. Total RNA was isolated from each of the frozen whole-bodies of beetles. mRNA was purified from the total RNA using oligo-dT attached magnetic beads. Following mRNA fragmentation, cDNA synthesis, end repair, 3' adenylation, adaptor ligation, and PCR enrichment, four RNA sequencing (RNA-Seq) libraries were constructed using the Illumina TruSeq RNA sample preparation kit. The raw sequence data were filtered by removing adaptors, low-quality reads, and ambiguous reads. We ultimately obtained 55.4, 43.4, 38.6, and 36.7 million clean reads for *Rhagophthalmus* sp., *A. circumdata*, *A. ficta*, and *P. pectoralis*, respectively (Table 1). These clean reads were separately assembled into 44883, 57254, 71424, and 80017 contigs, in which 15418, 24275, 31520, and 31356 unigenes (i.e. unique putative genes) were derived, respectively (Table 1). It appears that the transcriptome data in greater size tend to have less assembled contigs or unigenes. For example, the largest data (*Rhagophthalmus* sp.) contains the least contigs, while the smallest data (*P. pectoralis*) consists of the most contigs (Table 1). Details of clean reads, assembled contigs and unigenes of the four transcriptomes generated in this study were given in Table 1. In addition, we compared the 13 PCGs in *A. ficta* mitogenome identified from the Illumina-based transcriptome assembly with the same genes obtained from the traditional Sanger sequencing for another individual of the same species, and found that a total of 99.3% sequences from both sequencing approaches are identical. This finding suggests that our transcriptome assemblies are reliable and our sequencing coverages are acceptable.

Functional annotation of unigenes. A total of 10345 (percentage of all unigenes in a species: 67.10%), 15352 (63.24%), 17480 (55.46%), and 15342 (48.93%) unigenes of *Rhagophthalmus* sp., *A. circumdata*, *A. ficta*, and *P. pectoralis* were respectively annotated by at least one of the publicly available databases: Swiss-prot, NCBI non-redundant protein (NR), Clusters of Orthologous Groups of proteins (COG), and Gene Ontology (GO) (Table 2, Supplementary Dataset S1). We used all unigenes as query and ran the TBLASTX program to search against the Swiss-prot, NR, COG, and GO databases with an E-value of 1.0E-5. For example, we identified 10322 best blast hits in the NR database when annotating the transcriptome of *Rhagophthalmus* sp., comprising 66.95% of all unigenes (Table 2); while only 4845 best hits were detected in the GO database for the same species, covering 31.42% of all unigenes (Table 2). Indeed, the NR database generally annotated the highest percentages of

Database	<i>Rhagophthalmus</i> sp.	<i>A. circumdata</i>	<i>A. ficta</i>	<i>P. pectoralis</i>
Swiss-prot	8,447 (54.79%)	11,171 (46.02%)	12,042 (38.20%)	11,263 (35.92%)
NR	10,322 (66.95%)	15,315 (63.09%)	17,432 (55.30%)	15,304 (48.81%)
GO	4,845 (31.42%)	6,403 (26.38%)	7,235 (22.95%)	6,428 (20.50%)
COG	9,027 (58.55%)	12,197 (52.25%)	13,357 (42.38%)	12,415 (39.59%)
Total	10,345 (67.10%)	15,352 (63.24%)	17,480 (55.46%)	15,342 (48.93%)

Table 2. Summary of unigene annotations for the four transcriptomes. Percentages of all unigenes in a given species were shown in parentheses. Abbreviations: NR, NCBI non-redundant proteins; COG, Clusters of Orthologous Groups of proteins; GO, Gene Ontology.

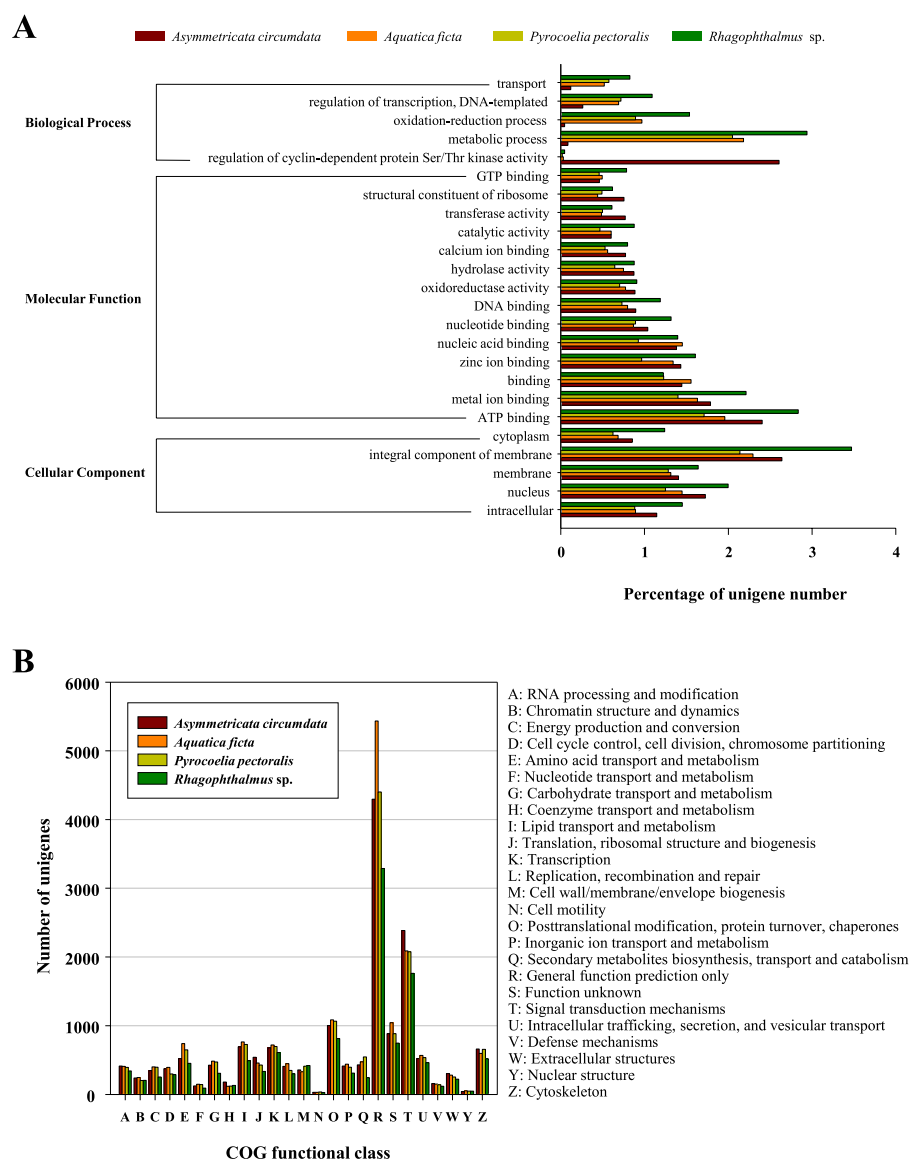


Figure 1. Functional classification of unigenes derived from the four transcriptomes based on the GO (Gene Ontology) annotation (A) and the COG (Cluster of Orthologous Groups) annotation (B).

unigenes, whereas the GO database typically annotated the lowest (Table 2). Overall, *Rhagophthalmus* sp. has a higher percentage of unigenes being annotated compared to the other three beetles (Table 2), possibly because a greater size of sequencing data can give a better assembly for the *Rhagophthalmus* beetle (Table 1). In addition, the distributions of both GO terms and COG classes were similar among the four transcriptomes, except *A. circumdata* showing a unique expansion in a single GO term (Fig. 1).

Gene	Direction	<i>Aquatica wuhana</i>			<i>Aquatica ficta</i>			<i>Lamprigera yunnana</i>			
		From	To	Start/stop codon	From	To	Start/stop codon	From	To	Start/stop codon	
<i>tRNA^{Ile}</i>	Forward	1	64	ATA/TAA	1	63	ATA/TAG			n.a./TAA	
<i>tRNA^{Gln}</i>	F	62	130		61	129					
<i>tRNA^{Met}</i>	Reverse	130	195		129	194					
<i>ND2</i>	F	196	1209		195	1208			1*		878
<i>tRNA^{Tyr}</i>	F	1211	1275		1210	1274			882		942
<i>tRNA^{Cys}</i>	R	1342	1404		1340	1402			1143		1203
<i>tRNA^{Tyr}</i>	R	1404	1466		1402	1465			1203		1265
<i>COI</i>	F	1438	3003	ATT/TAA	1458	3002	ATT/TAA	1237	2802	ATT/TAA	
<i>tRNA^{Leu}</i>	F	2999	3062		2988	3061			2798		2861
<i>COII</i>	F	3064	3742	ATG/T+tRNA	3063	3741	ATG/T+tRNA	2835	3540	ATA/T+tRNA	
<i>tRNA^{Lys}</i>	F	3743	3813		3742	3812			3541		3610
<i>tRNA^{Asp}</i>	F	3813	3876		3812	3874		3610	3675		
<i>ATP8</i>	F	3877	4032	ATT/TAA	3875	4030	ATT/TAA	3685	3828	ATA/TAA	
<i>ATP6</i>	F	4026	4700	ATG/TAA	4024	4698	ATG/TAA	3822	4485	ATG/T+tRNA	
<i>COIII</i>	F	4700	5483	ATG/T+tRNA	4698	5481	ATG/T+tRNA	4486	5266	ATG/T+tRNA	
<i>tRNA^{Gly}</i>	F	5484	5546		5482	5545			5267		5328
<i>ND3</i>	F	5547	5900	ATA/TAG	5546	5899	ATT/TAG	5335	5682	ATA/TAG	
<i>tRNA^{Ala}</i>	F	5899	5962		5898	5961			5681		5744
<i>tRNA^{Arg}</i>	F	5962	6026		5961	6025			5743		5805
<i>tRNA^{Asn}</i>	F	6026	6090		6025	6090			5804		5865
<i>tRNA^{Ser}</i>	F	6091	6157		6091	6157			5858		5920
<i>tRNA^{Glu}</i>	F	6158	6220		6158	6221			5921		5983
<i>tRNA^{Phe}</i>	R	6219	6281		6220	6283			5982		6041
<i>ND5</i>	R	6282	7989	ATT/T+tRNA	6284	7994	ATA/T+tRNA	6039	7749	ATT/T+tRNA	
<i>tRNA^{His}</i>	R	7990	8052		7992	8054			7750		7810
<i>ND4</i>	R	8053	9379	ATG/T+tRNA	8055	9381	ATG/T+tRNA	7810	9135	ATG/TAG	
<i>ND4L</i>	R	9373	9663	ATG/TAA	9375	9665	ATG/TAA	9129	9416	ATG/TAA	
<i>tRNA^{Thr}</i>	F	9665	9727		9667	9728			9418		9480
<i>tRNA^{Pro}</i>	R	9727	9790		9729	9794			9481		9544
<i>ND6</i>	F	9792	10277	ATA/TAA	9796	10284	ATA/TAA	9546	10049	ATT/TAA	
<i>CYTB</i>	F	10277	11410	ATG/TAG	10284	11417	ATG/TAG	10050	11177	ATG/TAG	
<i>tRNA^{Ser}</i>	F	11409	11474		11416	11481			11176		11240
<i>ND1</i>	R	11479	12437	ATT/TAG	11499	12443	ATT/TAG	11257	12207	ATG/TAG	
<i>tRNA^{Leu}</i>	R	12445	12506		12451	12513			12209		12269
<i>16SrRNA</i>	R	12507	13773		12514	13776			12270		13525
<i>tRNA^{Val}</i>	R	13774	13842		13777	13846			13525		13592
<i>12SrRNA</i>	R	13843	14588		13847	14596			13588		14138*
A+T-rich region		14589	16186		14597	16836			n.a.		

Table 3. Annotations of the three newly sequenced mitochondrial genomes. Incomplete sequenced region was indicated with an asterisk. Abbreviation: n.a., not available (due to incomplete sequencing).

Ortholog identification. Genes from the reference genome of the red flour beetle *Tribolium castaneum* were respectively compared with all unigenes from each of the four transcriptome assemblies using the TBLASTX searches²⁷. By examining reciprocal (or bi-directional) best blast hits, putative orthologs were identified. If the BLAST score ratio of the second best-hit to the first best-hit is greater than 0.8, we excluded the putative ortholog for further analysis, because the candidate ortholog is possibly a paralog due to the high level of sequence similarity. The resulting putative orthologs range from 6992 to 7612 (Table 2). Each ortholog from the red flour beetle and the four beetles with transcriptome data was aligned by PRANK version 100802²⁸, and poorly aligned positions and divergent regions in alignments were filtered by GBLOCKS version 0.91b²⁹. After discarding the alignments with an aligned region shorter than 100 nt (nucleotides), we obtained 4325 putative orthologs with high-quality alignments for subsequent analyses.

Mitogenome sequencing. Using the traditional Sanger sequencing approach, we sequenced two complete mitogenomes (*A. ficta* and *A. wuhana*) and one nearly complete mitogenome (*L. yunnana*) (Table 3). We were unable to sequence the region between *ND2* and *12S rRNA* in *L. yunnana* despite multiple attempts (Table 3), possibly because this region contains the A+T-rich region which may pose technical issues in sequencing³⁰. All the 13 protein coding genes (PCGs) from each of the three mitogenomes were complete except the *ND2* in

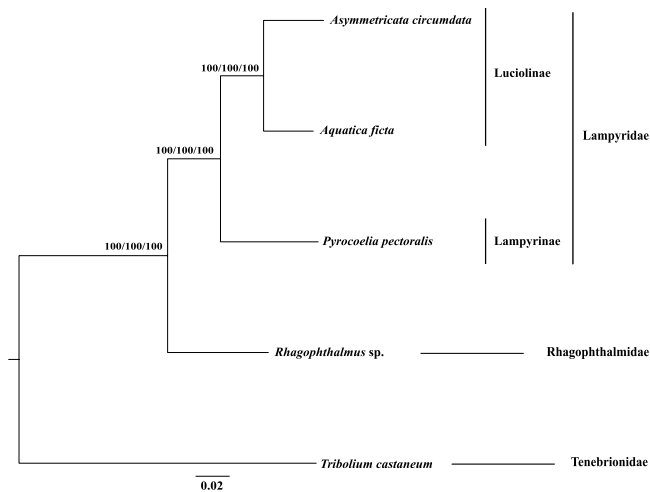


Figure 2. Phylogenetic relationships between Rhagophthalmidae and Lampyridae inferred from the concatenated 4325 nuclear gene markers with the third codon positions removed, *Tribolium castaneum* was chosen to be the outgroup. Numbers at each node are the ML bootstrap values/Bayesian posterior probabilities/MP-EST bootstrap values, shown as percentages.

L. yunnana lacking a segment near its 5' end (Table 3). All the 13 PCGs, 2 rRNAs, 22 transfer RNAs (tRNAs) and 1 A-T-rich region common to the vast majority of animal mitogenomes³¹ were identified in the three newly sequenced mitogenomes, except for the three tRNAs (tRNA^{Trp}, tRNA^{Cys}, tRNA^{Tyr}) and the A-T-rich region in *L. yunnana* (Table 3). Arrangements and orientations of all genes in the three mitogenomes are identical to other beetles^{25,32–34}. Similar to other firefly beetles, all the PCGs employed traditional mitochondrial start codons ATN and terminated with TAA, TAG or single T (Table 3).

Phylogenetic analysis based on nuclear gene sequences. Phylogenetic analysis was first undertaken with transcriptome-derived nuclear gene sequences. Given that Lampyridae was divided into two monophyletic clades, with one clade consisting of Lampyrinae and the other clade comprising Cyphonocerinae, Ototetrinae, Luciolinae¹³, we selected one species *P. pectoralis* (Lampyrinae) from the first clade and two species *A. ficta* (Luciolinae) and *A. circumdata* (Luciolinae) from the second clade to perform whole-body transcriptome sequencing, in addition to the species *Rhagophthalmus* sp. The resulting transcriptome data were assembled and annotated, and a total of 4325 putative orthologs were identified after careful filtering of sequencing artifacts. A data set is considered to be mutationally saturated when multiple substitutions are dominating. Substitution saturation decreases phylogenetic information, and a highly saturated data set will produce an incorrect phylogeny²⁰. To assess the impact of substitution saturation, we randomly selected 100 nuclear orthologs and performed substitution saturation tests. Results showed that the third codon sites of most nuclear orthologs have undergone substantial substitution saturation (Supplementary Table S2), we thus only used the first and second codon sites of the 4325 nuclear orthologs (Supplementary Dataset S2) for subsequent phylogenetic analysis with the concatenation method³⁵. Based on the 1854774-nt concatenated alignment of nuclear orthologs from the four species with transcriptome data and the red flour beetle with reference genome data, we used both Maximum Likelihood (ML) and Bayesian Inference (BI) approaches to reconstruct phylogenetic trees. Both maximum likelihood bootstrap values and Bayesian inference posterior probability values (shown as percentages) were 100% at all nodes of the phylogenetic tree (Fig. 2). Specifically, both species from Luciolinae were first clustered into a sister clade, which next grouped with the single species from Lampyrinae (Fig. 2). All the three firefly beetles (Lampyridae) formed a monophyletic group, and the single species from Rhagophthalmidae appeared to locate outside of Lampyridae (Fig. 2).

To reduce the influence of gene tree heterogeneity, we also undertook phylogenetic analysis with the coalescent model using the MP-EST method³⁶. Based on the alignment of each orthologous gene with all three codon positions, the ML gene trees with 100 bootstrap replicates were respectively reconstructed by PhyML version 3.0³⁷ using the best-fit models generated by jModelTest version 2.1.4³⁸. Because poorly supported individual gene trees could reduce phylogenetic signals in inferring species tree^{39,40}, only 2,555 genes with average bootstrap support values greater than 70% were used for MP-EST tree reconstruction. Results showed that the tree topology and bootstrap values of the MP-EST tree using the coalescent method were identical to those of the ML and BI trees reconstructed using the concatenation method (Fig. 2).

For comparison, we additionally reconstructed a phylogeny using a bioluminescence related gene. Because bioluminescence in beetles is produced by catalyzing the oxidation of luciferin by luciferase⁴¹, the luciferase gene is preferable for this analysis. We used the luciferase gene of *Lampyris noctiluca* (Genbank accession: X89479) as the query to search against the assembled transcripts of each transcriptome using TBLASTN. The best hit from each transcriptome was selected as a candidate, and candidate luciferase genes of the four beetles were searched against NR (NCBI non-redundant proteins) to ensure accuracy in gene identification. Taking the luciferase gene of *Tribolium castaneum* as the outgroup, we generated an alignment using the five luciferase genes, and

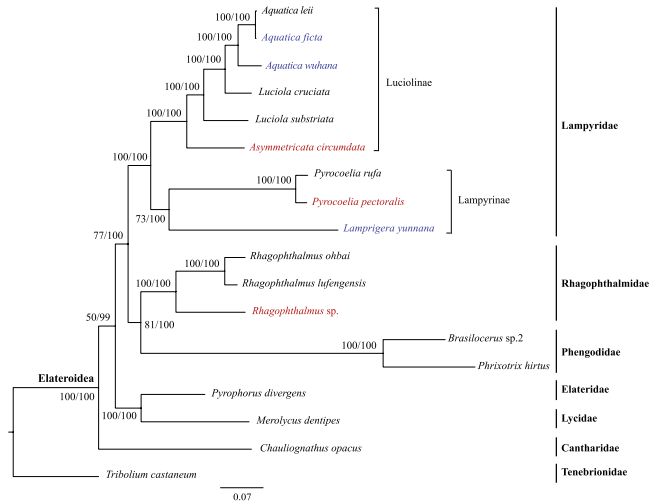


Figure 3. Phylogenetic relationships between Rhagophthalmidae and its related beetle families inferred from the concatenated 13 mitochondrial protein coding genes with the third codon positions removed. Numbers at nodes are the ML bootstrap values/Bayesian posterior probabilities, shown as percentages. Species shown in blue have mitochondrial genes generated from Sanger sequencing, while species in red have mitochondrial genes identified from Illumina-sequenced transcriptomes.

phylogenetic trees were built using similar approaches as described elsewhere^{19,42–47}. The resulting tree topologies (Supplementary Fig. S1) were identical to those inferred from 4325 genes (Fig. 2).

Phylogenetic analysis based on mitochondrial gene sequences. Phylogenetic analysis was also conducted with mitochondrial gene sequences. Firstly, we retrieved published mitogenomes from 12 species of related beetles: *Rhagophthalmus ohbai* (Rhagophthalmidae), *Rhagophthalmus lufengensis* (Rhagophthalmidae), *Pyrocoelia rufa* (Lampyridae), *Luciola cruciata* (Lampyridae), *Luciola substriata* (Lampyridae), *Aquatica leii* (Lampyridae), *Brasilocerus sp.2* (Phengodidae), *Phrixotrix hirtus* (Phengodidae), *Pyrophorus divergens* (Elateridae), *Merolycus dentipes* (Lycidae), *Chauliognathus opacus* (Cantharidae), and *Tribolium castaneum* (Tenebrionidae) (Fig. 3). Secondly, we newly determined mitogenomes with the traditional Sanger sequencing from three firefly beetles: *A. ficta* (Lampyridae), *A. wuhana* (Lampyridae) and *L. yunmana* (Lampyridae) (Table 3). Thirdly, we identified mitochondrial gene sequences from our newly determined transcriptome assemblies of four beetles: *Rhagophthalmus sp.* (Rhagophthalmidae), *A. ficta* (Lampyridae), *A. circumdata* (Lampyridae), and *P. pectoralis* (Lampyridae) (Supplementary Table S1). The firefly *A. ficta* has sequence data from both Sanger and Illumina sequencing, but the transcriptome-derived mitogenome sequence is incomplete (Supplementary Table S1), we thus selected the Sanger-based mitogenome sequence (Fig. 3 and Table 3) for further analysis. In total, our dataset of mitogenomes contained 18 beetles. In addition to the outgroup species *T. castaneum*, six families of beetles were included: three rhagophthalmids, nine lampyrids, two phengodids, one elaterid, one lycid, and one cantharid (Fig. 3).

The 13 mitochondrial PCGs from each of the 18 beetle species were aligned and concatenated. Substitution saturation analysis revealed that the third codon sites of all examined mitochondrial PCGs have experienced substantial substitution saturation, while the first and second codon sites have not (Supplementary Table S3), we thus just used the first and second codon sites of the 13 mitochondrial genes for subsequent phylogenetic analysis with the concatenation method (Supplementary Dataset S3). The concatenated alignment of the 13 mitochondrial PCGs is 7382 nt in length. Both ML and BI phylogenetic analyses recovered identical phylogenetic trees (Fig. 3). Our trees showed that *Rhagophthalmus sp.* clustered with two other rhagophthalmids with published mitogenomes, and the three rhagophthalmids formed a monophyletic group with a strong support of 100% in both ML and BI analyses (Fig. 3), confirming that our sampled *Rhagophthalmus* beetle indeed belongs to Rhagophthalmidae phylogenetically. As depicted from the trees, Rhagophthalmidae was first allied with Phengodidae as sister groups, and Rhagophthalmidae and Phengodidae formed a monophyletic group with Lampyridae (Fig. 3), confirming that Rhagophthalmidae is not a subgroup within Lampyridae. The long branches in Phengodidae species would not affect our phylogenetic analysis, because we recovered similar trees after removing the two phengodids (Supplementary Fig. S2). We did not perform the coalescent method in the phylogenetic analysis based on mitogenomes, because mitochondrial genes exhibit limited incongruence and the impact of mitochondrial gene tree heterogeneity has been believed to be minimal⁴⁸.

Discussion

In this study, we present the first whole transcriptome shotgun sequencing data in Rhagophthalmidae and Lampyridae species using massive parallel mRNA sequencing (RNA-seq), providing valuable genome resources for studying the evolution of intriguing traits in these beetles such as bioluminescence. We also newly determined two complete and one nearly complete mitogenome sequences in Lampyridae species. Using various phylogenetic approaches, our transcriptome and mitogenome data unambiguously demonstrate Rhagophthalmidae being

distinct from Lampyridae and reject previous hypotheses supporting Rhagophthalmidae as a subgroup within Lampyridae. Our molecular phylogenetic study supports the familial status of Rhagophthalmidae.

The phylogenetic position of Rhagophthalmidae was controversial due to conflicting results inferred from molecular and morphological data. For example, Rhagophthalmidae had been assigned to the family Phengodidae inferred from morphological data¹¹; Rhagophthalmidae had been considered to be a subfamily or genus in the family Lampyridae based on morphological, embryological and molecular evidence^{8,12,14,15}. However, the overwhelming molecular evidence has consistently recognized the distinctiveness of Rhagophthalmidae, which appeared to be different from both Lampyridae and Phengodidae^{7,9,13,17,22–24,26,49,50}. As a result, the familial status of Rhagophthalmidae has been well established. In this study, we used 4325 nuclear genes derived from transcriptome data and all 13 mitochondrial protein coding genes to conduct phylogenetic inference. Our analyses based on various analytical approaches and several datasets consistently recognized the distinctiveness of Rhagophthalmidae (Figs 2 and 3). Although our study did not involve multiple individuals or populations from each species, intraspecific genetic variation should not impact the phylogenetic resolution at the family or subfamily level. Although our samples are limited due to the lack of genetic material, we included taxa that may represent basal lineages. More importantly, our phylogenetic resolution of Rhagophthalmidae remained consistent after using different analytical approaches and multiple data sets (Figs 2 and 3). In support of our findings, morphological evidence has identified the distinctiveness of Rhagophthalmidae from Lampyridae and Phengodidae^{10,16}, and molecular evidence has revealed that Rhagophthalmidae is a sister group to Phengodidae^{9,13}. In contrast to our findings, two molecular studies argued that Rhagophthalmidae should be placed within Lampyridae^{8,15}. However, this argument was inferred from a short *16S rRNA* gene segment with a length of 506-nt^{8,15} and cannot be supported by multi-locus data sets^{9,13}. Although broad sampling of taxa is important in molecular phylogenetic studies^{8,15}, sufficient number of gene loci is also required to build a reliable phylogenetic tree^{39,51}. According to our mitogenome analysis, the three luminescent beetle families (Rhagophthalmidae, Lampyridae, and Phengodidae) formed a monophyletic group (Fig. 3), suggesting a single origin of bioluminescence in the common ancestor of these beetles. This finding is consistent with a recent phylogenetic analysis based on mitogenomes⁹ but inconsistent with another study based on a combination of mitochondrial and nuclear genes¹³. The contrasting results between mitochondrial and nuclear genes call for an in-depth study by adding more taxa and more genome-scale data, which will help reconstruct a reliable phylogeny and make a conclusive inference on the evolution of bioluminescence in beetles.

Phylogenomics refers to phylogenetic analysis involving genome-scale data. Phylogenomics has been believed to outcompete single-gene phylogenetics, which frequently yielded conflicting results caused by stochastic errors from small-scale data sets^{20,51}. However, systematic errors are still present after adding more data⁵². There are three major challenges that could generate strong incongruence in phylogenomic analysis: substitution saturation, nucleotide compositional bias, and different tree reconstruction methods²⁰. We took the following steps to overcome these challenges. First, we undertook substitution saturation tests and removed the third codon positions of all genes that may have undergone substantial substitution saturation. Both nuclear and mitochondrial gene data sets with the first and second codon positions consistently supported Rhagophthalmidae to be an independent group that is distinct from Lampyridae (Figs 2 and 3). Second, we used the deduced protein sequences to reduce nucleotide compositional bias. Our result showed that the BI tree topologies inferred from protein sequences (Supplementary Fig. S3) were identical to those inferred from nucleotide sequences (Figs 2 and 3), suggesting that nucleotide compositional bias is not a major factor affecting our phylogenetic analysis. Third, we used both ML and BI methods with nuclear and mitochondrial gene data sets, and identified no incongruence between the two tree reconstruction methods (Figs 2 and 3). In addition, given the prevalence of gene tree heterogeneity^{53–55}, we also conducted phylogenetic analysis with the coalescent method, which has been proved to generate accurate and congruent phylogenies in the presence of heterogeneous gene trees^{56,57}. The coalescent method implemented in the MP-EST program³⁶ was applied to our transcriptome-derived nuclear gene data set of 2555 putative orthologous genes, and yielded a phylogenetic tree identical to the concatenation-based ML and BI trees (Fig. 2). Indeed, after examining the proportions of 14 possible topologies inferred from the 2555 orthologous genes, we found that most genes (75.7%) supported the topology shown in Fig. 2 and the average proportion of other 13 topologies is 1.9% (Fig. 4 and Supplementary Table S4), suggesting a low level of gene tree heterogeneity among beetles studied in this work. This study proved usefulness of our genome-scale data in phylogenetic analysis, and will help to illuminate the origin and evolution of bioluminescence in Lampyridae and other luminescent beetles.

Materials and Methods

Ethics statement. All beetle species used in this study were sampled in the field. No specific permits were required, and no endangered or protected species were involved. All experiments on the beetles were conformed to the rules and guidelines on animal experimentation in China.

Taxon sampling and DNA extraction. Live larvae or adults of beetles studied here were sampled at five locations in China (Supplementary Table S5). All these individuals were stored at -80°C after freezing in liquid nitrogen. For mitogenome sequencing, genomic DNAs of the three species (*A. ficta*, *A. wuhana* and *L. yunmana*) were isolated from thoracic muscles of adult individuals with the Qiagen DNeasy kits.

Whole-body transcriptome sequencing (RNA-seq). Following the manufacturer's protocol, total RNAs were isolated using Trizol (Invitrogen) from whole bodies of four individuals: one adult individual of *Rhagophthalmus* sp. and three larva individuals of firefly beetles (*A. ficta*, *A. circumdata*, and *P. pectoralis*). Four paired-end libraries with an insert size of approximately 200 bp were constructed using the Illumina Truseq RNA sample prep kit according to the manufacturer's protocol. Details of constructing RNA-seq libraries were

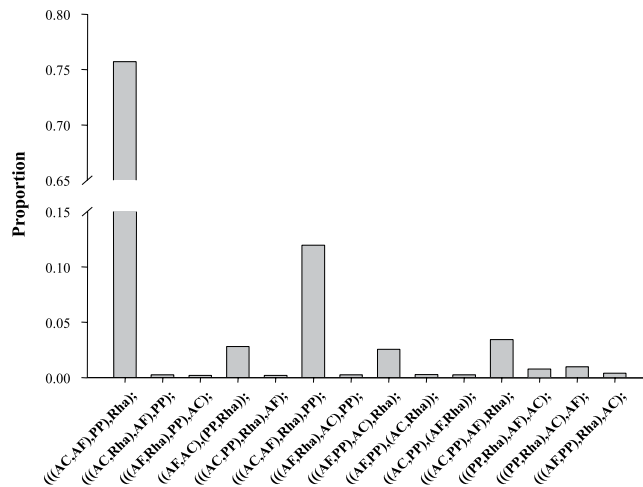


Figure 4. A histogram depicting the proportions of 14 tree topologies with average bootstrap values above 70%, inferred from 2555 orthologous genes in the MP-EST analysis. Rha, *Rhagophthalmus* sp.; AC, *Asymmetricata circumdata*; AF, *Aquatica ficta*; PP, *Pyrocoelia pectoralis*.

previously described¹⁹. All libraries were sequenced commercially to generate paired-end reads of average length 101-bp on the Illumina HiSeq 2000 sequencing platform⁵⁸.

De novo assembly and unigene annotation. Sequence quality of each RNA-seq sample was assessed by FastQC version 10.1 (www.bioinformatics.bbsrc.ac.uk/projects/fastqc). Trimmomatic version 0.32⁵⁹ was used to trim out low quality sequences, ambiguous sequences and artificial sequences such as residual adaptors and Illumina specific sequences as described elsewhere⁶⁰. Four de novo transcriptome assemblies were constructed by the Trinity program⁶¹ with default settings. Because de novo transcriptome assemblies involve many contigs in a contig cluster, we required a minimum expression filter of one fragment per kilobase of exon per million fragments mapped (FPKM) and filtered the contigs with FPKM value smaller than one; the contig with the highest expression level in a contig cluster was selected to be a unigene for downstream analyses⁶⁰. The resulting unigenes were used for BLASTX searches²⁷ and unigene annotations based on the NR, Swiss-prot and COG databases, with an E-value cutoff of 1e-5. The GO annotation was conducted by Blast2GO software⁶² with default parameters using the NR blast results in XML format.

PCR amplifications and mitogenomes annotation. To amplify the three mitogenomes (*A.ficta*, *A.wuhana* and *L.yunnana*), dozens of primer pairs were designed according to published firefly mitogenome sequences^{32–34}. Details of PCR amplification and sequencing procedure were described in our previous study³³. We defined the 13 PCGs of the three species by multiple sequence alignments with related species using MEGA version 5.20⁶³. The tRNAs were identified by tRNAscan-SE⁶⁴. The published mitogenomes used here were downloaded from the GenBank database (<http://www.ncbi.nlm.nih.gov/genbank>) under accession numbers as follows: *Tribolium castaneum*, NC_003081; *Brasilocerus* sp.2., KJ938490; *Pyrophorus divergens*, NC_009964; *Chauliognathus opacus*, NC_013576; *Rhagophthalmus ohbai*, NC_010964; *Rhagophthalmus lufengensis*, NC_010969; *Pyrocoelia rufa*, NC_003970; *Luciola cruciata*, NC_022472; *Luciola substriata*, NC_027176; *Aquatica leii*, NC_025276; *Phrixotrix hirtus*, KM923891; *Merolycus dentipes*, HQ232815.

Ortholog identification and phylogenomic analysis. The cDNA sequences of the red flour beetle (*Tribolium castaneum* 3.0 Assembly) were downloaded from BeetleBase⁶⁵ (<http://www.Beetlebase.org>), and compared with the transcriptome-derived unigenes by reciprocal (or bi-directional) TBLASTX searches²⁷. We identified putative orthologs by examining reciprocal best blast hits. We discarded putative orthologs with a BLAST score ratio of the second best-hit to the first best-hit greater than 0.8, which can exclude potential paralogs in our phylogenetic analysis. All putative orthologs were aligned by PRANK version 100802²⁸, and poorly aligned positions and divergent regions were removed by GBLOCKS version 0.91b²⁹. In addition, the alignments with an aligned region shorter than 100-nt were discarded.

All the 13 mitochondrial PCGs and 100 randomly selected nuclear orthologs were used to examine the substitution saturation by DAMBE⁶⁶. According to the Akaike information criterion (AIC) and Bayesian information criterion (BIC)⁶⁷, we ran the jModelTest version 2.1.4 program³⁸ separately to select the best-fit models of nucleotide substitution for concatenated nuclear and mitochondrial gene alignments after removal of the saturated third codon positions. For the concatenated nuclear genes, RAXML version 7.2.6⁶⁸ was used to reconstruct the ML tree under the GTR+GAMMA model with 100 bootstrap replicates, and MrBayes version 3.2.6⁶⁹ was applied to reconstruct the BI tree using the recommended GTR+I+G model with 0.5 million generations^{45,47}. To reduce the impact of gene tree heterogeneity, we also undertook phylogenetic analysis with the coalescent model using the MP-EST method³⁶. For the concatenated 13 mitochondrial PCGs, the ML tree was reconstructed by RAXML version 7.2.6 with 1000 bootstrap replicates under recommended GTR+GAMMA model, and the concatenated sequence was partitioned by different genes to estimate and optimize individual α -shape parameters, GTR-rates,

and base frequencies for each gene. MrBayes version 3.2.6 was used to reconstruct the BI tree with one million generations and the concatenated sequence was partitioned according to different models: HKY+G model for *ND4L*, GTR+G model for *ND1* and *ATP6*, and GTR+I+G model for the other 10 mitochondrial genes.

In addition, we used deduced protein sequences in our phylogenomic analysis, aiming to reduce the impact of nucleotide compositional bias. Briefly, protein sequences of the nuclear and mitochondrial genes were deduced and aligned by MEGA version 5.20⁶³. ProtTest version 3.41⁷⁰ was applied to select the best-fit model of protein sequence evolution following the Akaike information criterion (AIC). The LG+I+G model was determined to provide the best fit to the concatenated protein sequences of nuclear genes, while the MtrRev+I+G model was selected for the concatenated protein sequences of mitochondrial genes. Phylogenetic trees were reconstructed by MrBayes version 3.2.6⁶⁹ with as many generations as required.

References

- Herring, P. J. Systematic distribution of bioluminescence in living organisms. *J. Biolumin. Chemilumin.* **1**, 147–163, doi:10.1002/bio.1170010303 (1987).
- Haddock, S. H., Moline, M. A. & Case, J. F. Bioluminescence in the sea. *Ann. Rev. Mar. Sci.* **2**, 443–493, doi:10.1146/annurev-marine-120308-081028 (2010).
- Poisson, J. Raphael Dubois, from pharmacy to bioluminescence. *Rev. Hist. Pharm. (Paris)* **58**, 51–56 (2010).
- Callaway, E. Glowing plants spark debate. *Nature* **498**, 15–16, doi:10.1038/498015a (2013).
- Lloyd, J. E. Insect bioluminescence, pp. 241–272, In: *Bioluminescence in Action*, Herring, P. (ed.) (Academic Press, 1978).
- Viviani, V. R. The origin, diversity, and structure function relationships of insect luciferases. *Cell. Mol. Life Sci.* **59**, 1833–1850, doi:10.1007/PL00012509 (2002).
- Bocakova, M., Bocak, L., Hunt, T., Teravainen, M. & Vogler, A. P. Molecular phylogenetics of Elateriformia (Coleoptera): evolution of bioluminescence and neoteny. *Cladistics* **23**, 477–496, doi:10.1111/j.1096-0031.2007.00164.x (2007).
- Li, X. Y., Yang, S., Xie, M. & Liang, X. C. Phylogeny of fireflies (Coleoptera: Lampyridae) inferred from mitochondrial 16S ribosomal DNA, with references to morphological and ethological traits. *Progr. Nat. Sci.* **16**, 817–826, doi:10.1080/10020070612330074 (2006).
- Amaral, D. T., Mitani, Y., Ohmiya, Y. & Viviani, V. R. Organization and comparative analysis of the mitochondrial genomes of bioluminescent Elateroidea (Coleoptera: Polyphaga). *Gene* **586**, 254–262, doi:10.1016/j.gene.2016.04.009 (2016).
- Branham, M. A. & Wenzel, J. W. The evolution of bioluminescence in cantharoids (Coleoptera: Elateroidea). *Fla. Entomol.* **84**, 565–586, doi:10.2307/3496389 (2001).
- Crowson, R. A. A review of the classification of Cantharoidea (Coleoptera) with the definition of two new families, Cneoglossidae and Omethidae. *Rev. Univ. Madrid* **21**, 35–77 (1972).
- Kobayashi, Y., Suzuiki, H. & Ohba, N. Formation of a spherical germ rudiment in the glow-worm *Rhagophthalmus ohbai* Wittmer (Coleoptera: Rhagophthalmidae) and its phylogenetic implications. *Proc. Arthropod. Embryol. Soc. Jpn.* **36**, 1–5 (2001).
- Kundrata, R., Bocakova, M. & Bocak, L. The comprehensive phylogeny of the superfamily Elateroidea (Coleoptera: Elateriformia). *Mol. Phylogenet. Evol.* **76**, 162–171, doi:10.1016/j.ympev.2014.03.012 (2014).
- McDermott, F. A. The taxonomy of the Lampyridae (Coleoptera). *Trans. Am. Entomol. Soc.* **90**, 1–72 (1964).
- Stanger-Hall, K. F., Lloyd, J. E. & Hillis, D. M. Phylogeny of North American fireflies (Coleoptera: Lampyridae): Implications for the evolution of light signals. *Mol. Phylogenet. Evol.* **45**, 33–49, doi:10.1016/j.ympev.2007.05.013 (2007).
- Wittmer, W. & Ohba, N. Neue Rhagophthalmidae (Coleoptera) aus China und benachbarten Ländern. *Jpn. Entomol.* **62**, 341–355 (1994).
- Kundrata, R. & Bocak, L. The phylogeny and limits of Elateridae (Insecta, Coleoptera): is there a common tendency of click beetles to soft-bodiedness and neoteny? *Zool. Scr.* **40**, 364–378, doi:10.1111/j.1463-6409.2011.00476.x (2011).
- Nei, M. & Kumar, S. *Molecular Evolution and Phylogenetics*. (Oxford University Press, 2000).
- Lin, G. H. *et al.* Transcriptome sequencing and phylogenomic resolution within Spalacidae (Rodentia). *BMC Genomics* **15**, 32, doi:10.1186/1471-2164-15-32 (2014).
- Jeffroy, O., Brinkmann, H., Delsuc, F. & Philippe, H. Phylogenomics: the beginning of incongruence? *Trends Genet.* **22**, 225–231, doi:10.1016/j.tig.2006.02.003 (2006).
- Kobayashi, Y., Suzuiki, H. & Ohba, N. Formation of a spherical germ rudiment in the glow-worm *Rhagophthalmus ohbai* Wittmer (Coleoptera: Rhagophthalmidae) and its phylogenetic implications. *Proc. Arthropod. Embryol. Soc. Jpn.* **36**, 1–5 (2001).
- Hunt, T. *et al.* A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science* **318**, 1913–1916, doi:10.1126/science.1146954 (2007).
- Timmermans, M. J. *et al.* Why barcode? High-throughput multiplex sequencing of mitochondrial genomes for molecular systematics. *Nucleic Acids Res.* **38**, e197–e197, doi:10.1093/nar/gkq807 (2010).
- Bocak, L. *et al.* Building the Coleoptera tree-of-life for >8000 species: composition of public DNA data and fit with Linnaean classification. *Syst. Entomol.* **39**, 97–110, doi:10.1111/syen.12037 (2014).
- Li, X., Ogoh, K., Ohba, N., Liang, X. C. & Ohmiya, Y. Mitochondrial genomes of two luminous beetles, *Rhagophthalmus lufengensis* and *R. ohbai* (Arthropoda, Insecta, Coleoptera). *Gene* **392**, 196–205, doi:10.1016/j.gene.2006.12.017 (2007).
- Timmermans, M. J. *et al.* Family-level sampling of mitochondrial genomes in Coleoptera: Compositional heterogeneity and phylogenetics. *Genome Biol. Evol.* **8**, 161–175, doi:10.1093/gbe/evv241 (2015).
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410, doi:10.1016/S0022-2836(05)80360-2 (1990).
- Loytynoja, A. & Goldman, N. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* **320**, 1632–1635, doi:10.1126/science.1158395 (2008).
- Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552, doi:10.1093/oxfordjournals.molbev.a026334 (2000).
- Vila, M. & Bjorklund, M. The utility of the neglected mitochondrial control region for evolutionary studies in lepidoptera (Insecta). *J. Mol. Evol.* **58**, 280–290, doi:10.1007/s00239-003-2550-2 (2004).
- Boore, J. L. Animal mitochondrial genomes. *Nucleic Acids Res* **27**, 1767–1780, doi:10.1093/nar/27.8.1767 (1999).
- Bae, J. S., Kim, I., Sohn, H. D. & Jin, B. R. The mitochondrial genome of the firefly, *Pyrocoelia rufa*: complete DNA sequence, genome organization, and phylogenetic analysis with other insects. *Mol. Phylogenet. Evol.* **32**, 978–985, doi:10.1016/j.ympev.2004.03.009 (2004).
- Jiao, H., Ding, M. & Zhao, H. Sequence and organization of complete mitochondrial genome of the firefly, *Aquatica leii* (Coleoptera: Lampyridae). *Mitochondrial DNA*, **0**, 1–2, doi:10.3109/19401736.2013.855746 (2013).
- Mu, F. J., Ao, L., Zhao, H. B. & Wang, K. Characterization of the complete mitochondrial genome of the firefly, *Luciola substriata* (Coleoptera: Lampyridae). *Mitochondrial DNA*, 1–3 (2015).
- William, J. & Ballard, O. Combining data in phylogenetic analysis. *Trends Ecol. Evol.* **11**, 334, doi:10.1016/0169-5347(96)81133-5 (1996).
- Liu, L., Yu, L. & Edwards, S. V. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evol. Biol.* **10**, 302, doi:10.1186/1471-2148-10-302 (2010).

37. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321, doi:10.1093/sysbio/syq010 (2010).
38. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9**, 772–772, doi:10.1038/nmeth.2109 (2012).
39. Song, S., Liu, L., Edwards, S. V. & Wu, S. Y. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proc. Natl. Acad. Sci. USA* **109**, 14942–14947, doi:10.1073/pnas.1211733109 (2012).
40. Zhong, B., Liu, L., Yan, Z. & Penny, D. Origin of land plants using the multispecies coalescent model. *Trends Plant Sci.* **18**, 492–495, doi:10.1016/j.tplants.2013.04.009 (2013).
41. Marques, S. M. & Esteves da Silva, J. C. Firefly bioluminescence: a mechanistic approach of luciferase catalyzed reactions. *IUBMB Life* **61**, 6–17, doi:10.1002/iub.134 (2009).
42. Hong, W. & Zhao, H. Vampire bats exhibit evolutionary reduction of bitter taste receptor genes common to other bats. *Proc. Biol. Sci.* **281**, 20141079, doi:10.1098/rspb.2014.1079 (2014).
43. Feng, P., Zheng, J. S., Rossiter, S. J., Wang, D. & Zhao, H. Massive losses of taste receptor genes in toothed and baleen whales. *Genome Biol. Evol.* **6**, 1254–1265, doi:10.1093/gbe/evu095 (2014).
44. Zhao, H. *et al.* The evolution of color vision in nocturnal mammals. *Proc. Natl. Acad. Sci. USA* **106**, 8980–8985, doi:10.1073/pnas.0813201106 (2009).
45. Wang, K. & Zhao, H. Birds generally carry a small repertoire of bitter taste receptor genes. *Genome Biol. Evol.* **7**, 2705–2715, doi:10.1093/gbe/evv180 (2015).
46. Zhang, Y., Jiao, Y., Jiao, H., Zhao, H. & Zhu, Y. X. Two-step functional innovation of the stem-cell factors WUS/WOX5 during plant evolution. *Mol. Biol. Evol.*, Forthcoming (2017).
47. Lu, Q., Wang, K., Lei, F., Yu, D. & Zhao, H. Penguins reduced olfactory receptor genes common to other waterbirds. *Sci. Rep.* **6**, 31671, doi:10.1038/srep31671 (2016).
48. Pollard, D. A., Iyer, V. N., Moses, A. M. & Eisen, M. B. Widespread discordance of gene trees with species tree in *Drosophila*: evidence for incomplete lineage sorting. *PLoS Genet.* **2**, e173, doi:10.1371/journal.pgen.0020173 (2006).
49. Mckenna, D. D. *et al.* The beetle tree of life reveals that Coleoptera survived end-Permian mass extinction to diversify during the Cretaceous terrestrial revolution. *Syst. Entomol.* **40**, 835–880, doi:10.1111/syen.12132 (2015).
50. Yuan, M. L. *et al.* High-level phylogeny of the Coleoptera inferred with mitochondrial genome sequences. *Mol. Phylogenet. Evol.* **104**, 99–111, doi:10.1016/j.ympev.2016.08.002 (2016).
51. Philippe, H., Delsuc, F., Brinkmann, H. & Lartillot, N. Phylogenomics. *Ann. Rev. Ecol. Evol. Syst.* **36**, 541–562, doi:10.1146/annurev.ecolsys.35.112202.130205 (2005).
52. Felsenstein, J. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27**, 401–410, doi:10.2307/2412923 (1978).
53. Edwards, S. V. Is a New and General Theory of Molecular Systematics Emerging? *Evolution* **63**, 1–19, doi:10.1111/j.1558-5646.2008.00549.x (2009).
54. Liu, L., Yu, L., Kubatko, L., Pearl, D. K. & Edwards, S. V. Coalescent methods for estimating phylogenetic trees. *Mol. Phylogenet. Evol.* **53**, 320–328, doi:10.1016/j.ympev.2009.05.033 (2009).
55. Belfiore, N. M., Liu, L. & Moritz, C. Multilocus phylogenetics of a rapid radiation in the genus *Thomomys* (Rodentia: Geomyidae). *Syst. Biol.* **57**, 294–310, doi:10.1080/10635150802044011 (2008).
56. Kubatko, L. S. & Degnan, J. H. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst. Biol.* **56**, 17–24, doi:10.1080/10635150601146041 (2007).
57. Degnan, J. H. & Rosenberg, N. A. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* **24**, 332–340, doi:10.1016/j.tree.2009.01.009 (2009).
58. Li, K. *et al.* Sympatric speciation revealed by genome-wide divergence in the blind mole rat *Spalax*. *Proc. Natl. Acad. Sci. USA* **112**, 11905–11910, doi:10.1073/pnas.1514896112 (2015).
59. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120, doi:10.1093/bioinformatics/btu170 (2014).
60. Harrison, P. W. *et al.* Sexual selection drives evolution and rapid turnover of male gene expression. *Proc. Natl. Acad. Sci. USA* **112**, 4393–4398, doi:10.1073/pnas.1501339112 (2015).
61. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–U130, doi:10.1038/nbt.1883 (2011).
62. Gotz, S. *et al.* High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **36**, 3420–3435, doi:10.1093/nar/gkn176 (2008).
63. Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739, doi:10.1093/molbev/msr121 (2011).
64. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964, doi:https://doi.org/10.1093/nar/25.5.0955 (1997).
65. Kim, H. S. *et al.* BeetleBase in 2010: revisions to provide comprehensive genomic information for *Tribolium castaneum*. *Nucleic Acids Res.* **38**, D437–D442, doi:10.1093/nar/gkp807 (2010).
66. Xia, X. H. DAMBE5: A comprehensive software package for data analysis in molecular biology and evolution. *Mol. Biol. Evol.* **30**, 1720–1728, doi:10.1093/molbev/mst064 (2013).
67. Posada, D. & Buckley, T. R. Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and Bayesian approaches over likelihood ratio tests. *Syst. Biol.* **53**, 793–808, doi:10.1080/10635150490522304 (2004).
68. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690, doi:10.1093/bioinformatics/btl446 (2006).
69. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755, doi:10.1093/bioinformatics/17.8.754 (2001).
70. Abascal, F., Zardoya, R. & Posada, D. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104–2105, doi:10.1093/bioinformatics/bti263 (2005).

Acknowledgements

Special thanks should go to Dr. Xinhua Fu (Huazhong Agricultural University) who kindly provided most beetle samples. We thank Dr. Chengquan Cao (Leshan Normal University) for collecting the sample of *Lamprigera yunnana*. We also thank Feng-Juan Mu, Yi Wang, and Liang Ao (Wuhan University) for technical assistance in the laboratory. This work was supported in part by the National Natural Science Foundation of China (91331115) and a Specialized Research Fund for the Doctoral Program of Higher Education (20130141110066).

Author Contributions

H.Z. conceived and designed the project. K.W., W.H. and H.J. performed the experiments. K.W. and H.Z. analyzed the data. H.Z. and K.W. wrote the manuscript. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at doi:[10.1038/s41598-017-01835-9](https://doi.org/10.1038/s41598-017-01835-9)

Competing Interests: The authors declare that they have no competing interests.

Accession Codes: Raw transcriptome data of the four luminescent beetles are available from the NCBI Sequence Read Archive under the accession number SRP082462. Newly determined mitogenome sequences of the three firefly beetles have been deposited to the GenBank under accession number KX758085-KX758087.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017