Genomic and experimental data provide new insights into luciferin biosynthesis and bioluminescence evolution in fireflies

Ru Zhang1,2,6, Jinwu He1,2,6, Zhiwei Dong2,6, Guichun Liu2,6, Yuan Yin1,6, Xinying Zhang3,6, Qi Li1,6, Yandong Ren2, Yongzhi Yang1, Wei Liu2, Xianqing Chen1, Wenhai Xia1, Kang Duan2, Fei Hao1, Zeshan Lin1, Jie Yang1, Zhou Chang2, Ruoping Zhao2, Wenting Wan1,6, Sihan Lu1, Yanqiong Peng4, Siqin Ge3*, Wen Wang1,2,5* & Xueyan Li2*

Fireflies are among the most charismatic insects for their spectacular bioluminescence, but the origin and evolution of bioluminescence remain elusive. Especially, the genic basis of luciferin (d-luciferin) biosynthesis and light patterns is largely unknown. Here, we present the high-quality reference genomes of two fireflies Lamprigera yunnana (1053 Mb) and Abscondita terminalis (501 Mb) with great differences in both morphology and luminous behavior. We sequenced the transcriptomes and proteomes of luminous organs of two species. We created the CRISPR/Cas9-induced mutants of Abdominal B gene without luminous organs in the larvae of A. terminalis and sequenced the transcriptomes of mutants and wild-types. Combining gene expression analyses with comparative genomics, we propose a more complete luciferin synthesis pathway, and confirm the convergent evolution of bioluminescence in insects. Using experiments, the function of the firefly acyl-CoA thioesterase (ACOT1) to convert L-luciferin to D-luciferin was validated for the first time. Comparisons of three-dimension reconstruction of luminous organs and their differentially expressed genes among two species suggest that two positive genes in the calcium signaling pathway and structural difference of luminous organs may play an important role in the evolution of flash pattern. Altogether, our results provide important resources for further exploring bioluminescence in insects.

Bioluminescence is a particularly intriguing phenomenon1, and its origin and evolution fascinate biologists since the time of Charles Darwin2. Fireflies (Lampyridae) are one of the best-known luminous organisms since the time of Aristotle1, and thus an important subject of scientific studies, especially related to their bioluminescent behavior and biochemistry. Together with other luminous beetles (Ragophthalmidae, Phengodidae and some Elateridae) in the same superfamily Elateroidea1, fireflies can produce light in the peroxisome of photocytes within the luminous organs with diverse morphology and location1 by a common oxidative mechanism of luciferin catalyzed by luciferases in the presence of ATP, O2 and Mg2+. Beetle luciferases have long been studied extensively in their sequence, structure and function4,5 and yielded numerous molecular, biomedical, pharmaceutical and bioanalytical applications5. Beetle luciferin, the substrate for bioluminescence reaction, appears to be conserved in structure among all luminous beetles, but is not found in non-luminous insects6, suggesting that its evolutionary origin may coincide with the origin of bioluminescence. Firefly bioluminescence first evolved as aposematic warning signal in larvae (glow)7 and later was co-opted as sexual signal in adults (glow, flash)7,8. Light on/off is controlled by the accessibility of O2 to peroxisome in photocytes, which is regulated by oxygen

1School of Ecology and Environment, Northwestern Polytechnical University, Xi’an 710072, Shaanxi, China. 2State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, Yunnan, China. 3Institute of Zoology, Chinese Academy of Sciences, Beijing, China. 4CAS Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla 666303, Yunnan, China. 5Center for Excellence in Animal Evolution and Genetics, Kunming 650223, Yunnan, China. 6These authors contributed equally: Ru Zhang, Jinwu He, Zhiwei Dong, Guichun Liu, Yuan Yin, Xinying Zhang and Qi Li. *email: gesq@ioz.ac.cn; wwang@mail.kiz.ac.cn; lxy@mail.kiz.ac.cn
nitrogen (NO) synthesis in tracheolar end cells induced by octopamine released from neural system through G-protein coupled receptor cAMP/PKA-Ca/Calmodulin signaling cascade\textsuperscript{8–11}. So far, the reference genomes of four luminous beetles (Lamyrinae (fireflies): 3; Elateridae (click beetles): 1) were reported in two separated articles\textsuperscript{12,13}. In one of them, the comparative genomics of two fireflies and one luminous click beetle support parallel origin of bioluminescence in beetles\textsuperscript{12}. However, the origin and evolution of luciferase genes and how bioluminescent light signal pattern (glow, flash) evolve in luminous beetles remain elusive. Most seriously, the genomic basis of luciferin biosynthesis is largely unknown.

\textit{Lamprigera yunnana} (Lamyrinae: incertae sedis) and \textit{Abscondita terminalis} (Lamyrinae: Luciolinae) display glow or flash signals, respectively, and also show great differences in both outer morphology and inner structure of their luminous organs (Fig. 1a–t; Supplementary Note 1, Figs. S1–S8, Videos S1–S4). Here, we present their high-quality reference genomes using single-molecule real-time (SMRT) sequencing technologies. A thorough investigation integrating multilevel data (including comparative genomics, proteomics and transcriptomics of luminous organs and their three-dimension reconstruction, functional verification of genes in vitro experiments, and CRISPR/Cas9 gene editing) provides new perspectives on luciferin biosynthesis, the origin and evolution of bioluminescence and light pattern.

\textbf{Results}

\textbf{Genome sequencing, assembly and annotation.} Using SMRT long reads, we assemble the high-quality reference genomes of two fireflies \textit{L. yunnana} (1053 Mb; contig N50: 3.51 Mb) and \textit{A. terminalis} (501 Mb; contig N50: 1.21 Mb) with high genome heterozygosity (Table 1; Supplementary Note 2, Tables S1–S4, Figs. S9–S10). Three evaluation methods (illumina reads mapping, RNA reads mapping and Benchmarking Universal Single-Copy Orthologs (BUSCO)) show the completeness and reliability of the two assemblies (Supplementary Note 2, Tables S5–S7). The assembled sizes are consistent with those estimated by k-mer analyses (Supplementary Table S2) and flow-cytometry\textsuperscript{14}. Among the assembled genomes of six phylogenetically related luminous beetles including four previously reported (three fireflies and one click beetle)\textsuperscript{12,13} and two fireflies sequenced in this study (Fig. 2a,b), \textit{L. yunnana} genome has the largest size (1053 Mb) and the highest percentage of repetitive elements (66.62%), while those of \textit{A. terminalis} (501 Mb; 36.54%) are similar to those of \textit{Photinus pyralis} (471 Mb; 47.70%) (Table 1; Supplementary Tables S8–S9). Comparative analyses of whole genomes among five fireflies (three previously reported\textsuperscript{12,13} and two sequenced here) indicate that genome size variation mainly results from the relative abundance of transposable elements (TEs), especially DNA transposons and long interspersed nuclear elements (LINEs), which are also two most abundant types of TEs among the genomes of all luminous beetles previously reported\textsuperscript{12,13} and sequenced in this study, and correlate in abundance with their host genome size (Table 1, Fig. 2c; Supplementary Table S9, Fig. S11). Combining de novo, homology-based and transcriptome-based methods, we predicted 19,443 and 21,024 genes in \textit{L. yunnana} and \textit{A. terminalis}, respectively (Supplementary Tables S10–S13). The gene structure features are similar to those of other fireflies (Supplementary Tables S11–S12, Fig. S12).

\textbf{Firefly phylogeny based on phylogenomic data.} We performed the phylogenomic analyses based on the 531 single-copy orthologous genes of Elateroidea and non-Elateroidea beetles plus fruit fly as an outgroup (Fig. 2a; Supplementary Note 4). Our results demonstrate that all six currently investigated luminous beetles (Lamyrinae (5) and Elateridae (1) in the families of Elateroidea\textsuperscript{14}) formed a clade (100% support), and diverged from non-Elateroidea beetles about 220-169 million years ago (Mya) (Fig. 2a), consistent with a previously reported phylogeny\textsuperscript{15} and the estimated divergent time of Elateroidea (182-152 Mya)\textsuperscript{16}. Considering that there are still no reference genomes available for other non-luminous families in Elateroidea, we constructed an additional mitogenomic phylogeny for 11 Elateroidea families (including luminous and non-luminous families) (Supplementary Table S14) to explore the phylogenetic distribution of bioluminescent within Elateroidea taxa. Our results (Fig. 2b) indicate that Lampyridae, together with other luminous families (Asian Rhagophthalmidae and South American Phengodidae), is a sister clade to world-wide Elateridae, a family with only some luminous species mainly in South America but recently also found by us in Asia\textsuperscript{17}. They corroborate with the phylogenies inferred from 95 nuclear protein-coding genes of beetles\textsuperscript{18} and from 13 protein-coding genes of mitogenomes (Supplementary Note 2, Tables S1–S4, Figs. S9–S10) and flow-cytometry\textsuperscript{14}. Among the assembled genomes of six phylogenetically related luminous beetles including four previously reported (three fireflies and one click beetle)\textsuperscript{12,13} and two fireflies sequenced in this study (Fig. 2a,b), \textit{L. yunnana} genome has the largest size (1053 Mb) and the highest percentage of repetitive elements (66.62%), while those of \textit{A. terminalis} (501 Mb; 36.54%) are similar to those of \textit{Photinus pyralis} (471 Mb; 47.70%) (Table 1; Supplementary Tables S8–S9). Comparative analyses of whole genomes among five fireflies (three previously reported\textsuperscript{12,13} and two sequenced here) indicate that genome size variation mainly results from the relative abundance of transposable elements (TEs), especially DNA transposons and long interspersed nuclear elements (LINEs), which are also two most abundant types of TEs among the genomes of all luminous beetles previously reported\textsuperscript{12,13} and sequenced in this study, and correlate in abundance with their host genome size (Table 1, Fig. 2c; Supplementary Table S9, Fig. S11). Combining de novo, homology-based and transcriptome-based methods, we predicted 19,443 and 21,024 genes in \textit{L. yunnana} and \textit{A. terminalis}, respectively (Supplementary Tables S10–S13). The gene structure features are similar to those of other fireflies (Supplementary Tables S11–S12, Fig. S12).

\textbf{Firefly phylogeny based on phylogenomic data.} We performed the phylogenomic analyses based on the 531 single-copy orthologous genes of Elateroidea and non-Elateroidea beetles plus fruit fly as an outgroup (Fig. 2a; Supplementary Note 4). Our results demonstrate that all six currently investigated luminous beetles (Lamyrinae (5) and Elateridae (1) in the families of Elateroidea\textsuperscript{14}) formed a clade (100% support), and diverged from non-Elateroidea beetles about 220-169 million years ago (Mya) (Fig. 2a), consistent with a previously reported phylogeny\textsuperscript{15} and the estimated divergent time of Elateroidea (182-152 Mya)\textsuperscript{16}. Considering that there are still no reference genomes available for other non-luminous families in Elateroidea, we constructed an additional mitogenomic phylogeny for 11 Elateroidea families (including luminous and non-luminous families) (Supplementary Table S14) to explore the phylogenetic distribution of bioluminescent within Elateroidea taxa. Our results (Fig. 2b) indicate that Lampyridae, together with other luminous families (Asian Rhagophthalmidae and South American Phengodidae), is a sister clade to world-wide Elateridae, a family with only some luminous species mainly in South America but recently also found by us in Asia\textsuperscript{17}. They corroborate with the phylogenies inferred from 95 nuclear protein-coding genes of beetles\textsuperscript{18} and from 13 protein-coding genes of mitogenomes and two nuclear ribosomal DNA (rDNA) (18S, 28S)\textsuperscript{19}, and with the beetle tree\textsuperscript{16} but differ from the phylogenies inferred from mitochondrial genes (16S, COI) and two nuclear rDNA (18S, 28S)\textsuperscript{15}. Although the so far reported phylogenies among Elateroidea families are still disputable, our data, together with previous findings\textsuperscript{3,15}, demonstrate a dispersedly phylogenetic distribution of bioluminescence in Elateroidea and even within Elateridae, suggesting a phenotypically convergent evolution of bioluminescence within beetles, as noted by Darwin\textsuperscript{2}. This phenomenon is similar to many recently scrutinized phenotypic traits such as feeding on poisonous milkweed for many insects, wing coloration patterns in butterflies and lateral plates in multiple sticklebacks\textsuperscript{20}.

Our phylogenetic (whole genome and mitogenome) analyses also demonstrate that \textit{L. yunnana}, is close to typical Luciolinae species (\textit{A. terminalis} and \textit{Aquatica lateralis}) with 100% support, and had diverged from Luciolinae about 56–103 Mya (Fig. 2a,b). This species has been placed originally in Lampyridae because of its similarities in morphology and luminous behavior to typical Lampyridae species (\textit{Pyrocoelia pectoralis} and \textit{P. pyralis})\textsuperscript{15}. We also compared 3125 single-copy orthologs among five fireflies, and the results show that \textit{L. yunnana} presented a higher average amino acid (AA) identity to Luciolinae (\textit{A. terminalis}: 77.44%; \textit{A. lateralis}: 78.95%) than Lampyridae (\textit{P. pyralis}: 74.39%; \textit{P. pectoralis}: 74.48%) (Fig. 2d), and that approximately its 65.96% genes are closer to those of \textit{A. terminalis} and \textit{A. lateralis} in sequence identity, while only 3.76% are closer to those of \textit{P. pyralis} and \textit{P. pectoralis} (Fig. 2e). The phylogenetic analysis of mitogenomic gene and rDNA genes\textsuperscript{15,20} and our comparison on morphology of \textit{Lamprigera} with that of a recently described fossil species in Luciolinae\textsuperscript{23} support the close relationship between \textit{Lamprigera} and Luciolinae. These combined data demonstrate that \textit{L. yunnana}
Figure 1. Adults of Lamprigera yunnana and Abscodita terminalis and their luminous organs. (a–d) Adults (a, male; b, female) and their luminous organs (dotted green regions) (c, male; d, female) of L. yunnana. (e–h) Adults (e, male; f, female) and their luminous organs (dotted green regions) (g, male; h, female) of A. terminalis. Scale bars (a–h): 1 mm. (i–l) Three-dimension reconstruction of luminous organs (blue) and tracheal system (red) of male and female L. yunnana (i, male; j, female) and A. terminalis (k, male; l, female). (m–p) Semi-thin section structures of luminous organs of male and female L. yunnana (m, male; n, female) and A. terminalis (o, male; p, female). (q–t) Transmission electron microscopic structures of luminous organs of male and female L. yunnana (q, male; r, female) and A. terminalis (s, male; t, female). PL, Photogenic layer; RL, Reflecting layer; T, Trachea.
Table 1. Comparison of reference genomes among beetles. *Contig †Chromosome and unplaced scaffolds §Scaffold §Predicted in this study using the same pipeline as in L. yunnana and A. terminalis. BUSCO: the Benchmarking Universal Single-Copy Ortholog.

| Family       | Species                      | Assembled size (Mb) | GC content (%) | N50 (Mb) | Repeat content (%) | DNA transposons (%) | Gene number | Completeness of BUSCO (%) | References |
|--------------|------------------------------|---------------------|----------------|----------|--------------------|--------------------|-------------|--------------------------|------------|
| Lampyridae   | Lampyrgena yunnana           | 1053                | 34.12          | 3.43*    | 66.62              | 34.91              | 19,443      | 98.67                    | This study |
|              | Abscondita terminalis        | 501                 | 31.39          | 1.18*    | 35.54              | 17.62              | 21,024      | 97.83                    | This study |
|              | Aquatica lateralis           | 902                 | 29.60*         | 0.69*    | 27.86*             | 14.84*             | 14,285      | 97.40                    |            |
|              | Photinus pennis              | 471                 | 36.41*         | 50.60*   | 47.70*             | 25.08*             | 15,773      | 97.20                    |            |
|              | Pyrocoelia pectoralis        | 760                 | 34.79          | 3.04*    | 47.21*             | 28.44*             | 23,092      | 97.83                    |            |
| Elateridae   | Ignelater luminosus          | 845                 | 33.422         | 0.12a    | 34.10              | 8.45               | 27,557      | 94.80                    |            |
|              | Agrilus planipennis          | 353                 | 35.99          | 1.11*    | 36.60*             | 3.20*              | 22,146      | 95.05§                    | GenBank    |
| Scarabaeidae | Onthophagus saxatus          | 267                 | 33.40          | 0.33*    | 48.08*             | 14.29*             | 21,668      | 99.46§                    | GenBank    |
| Cerambycidae | Anoplophora glabripennis     | 710                 | 33.39          | 0.66*    | 53.80*             | 6.84*              | 22,035      | 99.10§                    |            |
| Curculionidae| Dendroctonus ponderosae      | 252                 | 36.00          | 0.62*    | 23.00              | 0.31               | 13,088      | 95.35§                    |            |
| Tenebrionidae| Tribolium castaneum          | 166                 | 35.19          | 15.26†   | 4.75&              | 27.50              | 2.20        | 16,593                   |            |

has a closer phylogenetic relationship to Luciolinae than Lampyrinae, and thus L. yunnana should be a member of Luciolinae rather than a species in Lampyrinae.

Evolution of genes and gene families along Elateroidea. To explore the genomic basis of the origin and evolution of bioluminescence in insects, we performed a comparative genomics analysis among 21 species (six luminous beetles in Elateroidea and five non-luminous beetles in other five superfamilies (Coleoptera), nine representative species from five insect orders (Diptera, Lepidoptera, Hymenoptera, Hemiptera, Phthiraptera, Isoptera), and one Crustacea species (Branchiopoda)) (Fig. 3a; Supplementary Notes 5–6, Tables S15–S46, Figs. S13–S29, Data S1–S10). Analyses of gene family expansion and contraction show that 148 families are expanded in the ancestor of Elateridae-Lampyridae beetles (Elateroidea: currently all luminous beetle species are in this superfamily), of which gene ontology terms are significantly related to bioluminescence, peroxisome and catalytic activity, and KEGG is significantly related to the pathway of membrane transport (ABC transporters) and signal transduction (cAMP signaling pathway) (hyper test, corrected p < 0.01) (Fig. 3b; Supplementary Note 5, Tables S15–S30, Data S1–S4). The evolutionary analyses on genes among Elateroidea (only luminous taxa) and non-Elateroidea (all non-luminous taxa) (Supplementary Tables S31–S32, Figs. S13–S17, Data S5–S10) show that 190 orthologs are positively selected genes (PSGs) in the ancestor of Lampyridae-Elateridae (Elateroidea), which are mainly related to catalytic activity and ATP binding (Supplementary Data S5–S7). Specifically, these genes in calcium signaling (e.g. sarcoplasmic/endoplasmic reticulum calcium-transporting ATPase (SERCA), calreticulin) and in ATP binding cassette (ABC) transporter (i.e. ABC-D) were included. A thorough analysis of transcriptomes and proteomes of the adult luminous organs of L. yunnana calreticulin) and in ATP binding cassette (ABC) transporter (i.e. ABC-D) were included. A thorough analysis of transcriptomes and proteomes of the adult luminous organs of L. yunnana and A. terminalis (Fig. 3c; Supplementary Note 6, Tables S46, Figs. S18–S29, Data S11–S18) indicates that the highly expressed genes at both transcriptomic and proteomic levels in both species and sex are related to bioluminescence and ATP metabolic process. These results, combined with the dispersedly phylogenetic distribution of bioluminescence in Elateroidea and within Elateridae (Fig. 2b), suggest that convergent genetic evolution of these genes (families) in the luminous lineages of Elateroidea may contribute to the phenotypically convergent evolution of bioluminescence.

Origin and evolution of luciferase genes. To explore the origin of bioluminescence, we scrutinize the origin and evolution of luciferase among beetles (Supplementary Note 7, Figs. S30–S35, Data S19–S24) with more comprehensive methods than the previously reported13. Luciferase genes have been cloned from about 40 luminous beetles (Supplementary Data S19) and belong to acyl:CoA synthetase (ACS) superfamily with a close relationship to 4-coumarate:CoA ligase (4CL) family24. A thorough genome-wide investigation of ACS genes among beetles (six luminous beetles and five non-luminous beetles) (outgroup: Arabidopsis thaliana 4CL (Ath4CL1)) shows that a luciferase-like clade (including luciferase gene), together with its sister clade, 4-coumarate:CoA ligase, located at the terminus of the ACS gene tree and expanded greatly, almost occupying half of ACS genes in eight families (Supplementary Fig. S30). Further phylogenetic analysis on all luciferase-like and 4-coumarate:CoA ligase of beetles (Supplementary Fig. S31), together with those previously cloned luciferase homologs (Supplementary Data S19), demonstrated that except that of non-luminous Zophobas morio (ZopLL) belonging to 4-coumarate:CoA ligase, all other previously cloned luciferase-like homologs from non-luminous beetle Tenebrio molitor or from luminous beetles are luciferase-like genes. With our main aim to explore the
Origin of luciferases in luciferase-like family, we further focused our analysis on luciferase-like gene evolution (Fig. 4a). All above mentioned phylogenetic trees inferred from ACS, luciferase-like genes + 4-coumarate:CoA ligase or luciferase-like genes, show that an Elateroidea-specific luciferase-like clade evolved at the tree terminus, and within it, all luciferase genes in three phylogenetically related luminous taxa (i.e. Lampyridae, Rhagophthalmidae and Phengodidae) formed one terminal clade (marked by a red oval) with some of their paralogues at its base (marked by a brown oval) (Fig. 4a), which is sister to the Elateridae-luciferase + Elateridae-luciferase-like clade (marked by a purple oval) of Elateridae, a taxa including some luminous species. Our results are consistent with previously reported phylogeny of luciferases and their paralogues identified from the reference genomes of two fireflies and one luminous click beetle or non-luminous beetle genomes. We estimate that the ancestor of the luciferase gene in Lampyridae (plus Rhagophthalmidae, Phenagodidae) may have diverged around 205 Mya (Supplementary Figs. S32–S33), long before the divergence of Lampyridae and Elateridae inferred from
phylogenomic data (174-115 Mya). Elaterid luciferase gene, in contrast, evolved at a more recent time (~131 Mya) (Supplementary Fig. S3). Synteny analysis revealed the conserved syntonic blocks surrounding the luciferase locus across Lampyridae clades, which, however, is not syntenic to luciferase block in Elateridae (Fig. 4b). This suggests that luciferases in Lamyridae and Elateridae were evolved from different luciferase-like copies and different time. Amino acid sequence analysis indicates that all bioluminescent luciferases possess a pattern of “TSA/CSA/CCA” in a loop region between beta-sheets of N-terminal domain4,26 possibly interacting with luciferin27 and an overall amino acid identity of more than 47% to that of P. pyralis, suggesting that this amino acid sequence pattern played a key role in the bioluminescent function of beetle luciferase. All these data (phylogeny, divergence time, syntenic analysis) support that the bioluminescent function of luciferase genes was independently evolved in Lampyridae (plus Rhagophthalmidae and Phengodidae) and Elateridae, as proposed in a previous study13.

Luciferin biosynthesis revealed by multilevel data. We thoroughly investigated the genic basis of luciferin biosynthesis by integrating multilevel data including comparative genomics of luminous and non-luminous beetles, gene expression at both transcriptomic and proteomic levels in adult luminous organs of L. yunnana (Lyu) and A. terminalis (Ate) at both transcriptomic and proteomic levels. MO: luminous organ of male adult; FO: luminous organ of female adult.
In A. terminalis, the branch pathway of 1,4-hydroquinone (stored as arbutin) as the precursor of L-luciferin biosynthesis was previously proposed to be the precursor of L-luciferin biosynthesis here. Homogentisic acid (produced from firefly luciferin and sulfo-donor 3′-phosphoadenosine 5′-phosphosulfate (PAPS) produced from ATP and inorganic sulfate under the catalysis of PAPS synthase (PAPSS)) catalyzes the production of sulfoluciferin (a luciferin storage molecule, inactive for luciferase) from firefly luciferin and sulfo-donor 3′-phosphoadenosine 5′-phosphosulfate (PAPS) produced from ATP and inorganic sulfate under the catalysis of PAPS synthase (PAPSS). Our expression analysis showed that above mentioned enzymes involved with biosynthesis of D-luciferin and storage present a high expression at both transcriptomic and proteomic levels in the luminous organs of both species and sexes (Supplementary Figs. S52–S54). Thus, we retain the branch pathway of 1,4-hydroquinone (stored as arbutin) as the precursor of L-luciferin biosynthesis and can be further changed into L-luciferin in case of adding another L-cysteine. Our expression analysis showed a high expression (at both transcriptomic and proteomic levels) of some enzymes in tyrosine degradation (e.g., 4-hydroxyphenylpyruvate dioxygenase), hemocyanin (including polyphenol oxidase, hexamerin), 4-coumarate: CoA ligase and thiolase (ScpX, thiolase) in the luminous organs of both species and sexes (Fig. 5b–e; Supplementary Figs. S39, S52–S54, Data S25–S26). In mutants of A. terminalis generated by our gene-editing (Fig. 5f–j; Supplementary Note 8, Tables S52–S53, Figs. S60–S64, Data S30–S31), hexamerin was down-regulated while tyrosine hydroxylase (converting tyrosine to L-DOPA) was significantly up-regulated (Fig. 5j; Supplementary Data S30), suggesting that L-cysteine, one precursor of luciferin biosynthesis origins from cysteine anabolism. We identified the expression (transcriptomic and proteomic) of glucosidases in the luminous organs of both species and sexes (Supplementary Figs. S52–S54). Thus, we retain the branch pathway of 1,4-hydroquinone (stored as arbutin) as the precursor of L-luciferin biosynthesis here.

D-luciferin as the substrate of luciferase in firefly bioluminescence, is generated from the chirality transition of luciferin. The enzymes participating in conversion of L-luciferin to D-luciferin, including luciferase (LUC) for L-enantioselective thioesterification of L-luciferin and acyl-CoA thioesterase (ACOT) for hydrolysis, have been proposed. Moreover, a possible luciferin storage mechanism was proposed in fireflies that luciferin sulfotransferase catalyzes the production of sulfoluciferin (a luciferin storage molecule, inactive for luciferase) from firefly luciferin and sulfo-donor 3′-phosphoadenosine 5′-phosphosulfate (PAPS) produced from ATP and inorganic sulfate under the catalysis of PAPS synthase (PAPSS). Our expression analysis shows that above mentioned enzymes involved with biosynthesis of D-luciferin and storage present a high expression at both transcriptomic and proteomic levels in the luminous organs of both species and sexes (Fig. 5b–e; Supplementary Figs. S52–S54, Data S27). In the A. terminalis mutants (Fig. 5f–j; Supplementary Note 8, Tables S52–S53, Fig. S65, Data S30), luciferase and luciferin sulfotransferase were significantly down-regulated. The most noteworthy point is the role of the acyl-CoA thioesterases. Although a deracemizative luminescent system containing luciferase and acyl-CoA thioesterase in converting l-luciferin to d-luciferin, neither comprehensive genomic identification nor functional study on any insect acyl-CoA thioesterases was reported. Our phylogenomic investigation indicates that regardless of a great copy number variation (Supplementary Data S26), the acyl-CoA thioesterases of all investigated insects are belong to type-II ACOTs, and together mammal's (human and mouse) type-II ACOTs (63–67), they mainly cluster into two groups (Fig. 6a). One group (cluster-I) includes most mammalian type-II ACOTs (7, 9–12) at its base and some insect ACOTs at its terminus; and most of these insect ACOTs, like their closest sisters (i.e., mammalian mitochondrial acyl-CoA thioesterases (HomACOT9, MusACOT9-10)), exhibit eight similar gene sequence motifs and contain two 4HBT (4-hydroxybenzoyl-CoA thioesterase) domains (Fig. 6a). Interestingly, the luminous beetle-specific and single-copy syntenic orthologs (Fig. 6a) of those insect ACOTs show high expression in luminous organs at transcriptomic and proteomic levels (Supplementary Figs. S52–S54, Data S27). Another group (cluster-II), including mammalian ACOT13 (68–70) and multiple insect ACOT paralogs, show only two similar gene sequence motifs and one 4HBT domain, and only some of these luminous beetle ACOTs show expression in luminous organs at transcriptomic or proteomic levels (Fig. 6a; Supplementary Figs. S52–S54, Data S27). Additionally, some ACOTs from the two fireflies, together with some mammalian peroxisomal ACOTs (HomoACOT8 and MusACOT8), locate at the base (cluster-III) of all other ACOTs of insects and mammals, and show similar gene sequence motifs or protein domains to that of the fatty acyl-CoA thioesterase II (TESB, ACOTII) of E. coli in spite of no expression in luminous organs (Fig. 6a; Supplementary Fig. S53). Based on phylogenetic, gene sequence motifs and domain features, we selected three representative acyl-CoA thioesterases of A. terminalis from above mentioned three groups (AteACOT1: cluster-I, high expression at both transcriptomic and proteomic levels; AteACOT4: cluster-II, high expression...
and (Supplementary Fig. S54) were performed. Comacetyl-CoA thioesterase in insects. Importantly, this selected gene may promote the highest expressed acyl-CoA thioesterase (AteACOT1) (Fig. 5d,e; Supplementary Fig. S54) can efficiently convert l-luciferin to d-luciferin (Fig. 5k; Supplementary Fig. S47), which is the first verified the function of acyl-CoA thioesterase in insects.

Bioluminescent reaction occurs in the peroxisome in insects. However, the location of luciferin synthesis is still mysterious. A thorough whole-genome identification on peroxisome targeting signal (PTS) (incl. PTS1 or PTS2) and peroxisomal membrane proteins (two peroxin (Pex) genes, Pex5 and Pex14) were performed. Combined with the expression in luminous organs (Supplementary Note 7, Fig. S57, Data S28), our results suggest that peroxisomes are the function place of sterol carrier protein-X and luciferase. The positive selection on one member of the D subfamily of ATP-binding cassette gene family (ABC-D) in the ancestor of luminous beetles (Supplementary Note 7, Fig. S58, Data S6) and the high expression of its orthologs in L. yunnana (LY01293) and A. terminalis (LT01539) (Fig. 5b,c) suggest that, like human ABC-D genes (i.e. the import of long and branched chain acyl-CoA molecules into the peroxisome), this selected gene may promote the diffusion of cysteine (121 Da), 1,4-hydroquinone (108.09 Da) and 1,4-benzoquinone (108.09 Da) into peroxisome. These results provide the evidence that luciferin is synthesized in peroxisomes (Supplementary Note 7, Fig. S57, Data S27) of membrane channel Pmp2 protein (PMP22) in all luminous beetles, which can transfer metabolite of < 300 Da across the peroxisomal membrane, may contribute to function in luciferin deracemization (Fig. 6a). How acyl-CoA thioesterases are transferred into peroxisome in insects is still an open question because no peroxisomal targeting signals are identified in those fruit fly luciferases. We further explore the genetic causes of phenotypically convergent bioluminescence between Lampyridae and Elateridae, we assessed gene location (microsynteny) to confirm the orthology of major candidate genes (polyribonucleotide nucleotidyltransferase 1: PRNT1. (b) All bioluminescent luciferases show a pattern of “TSA/CSA/CCA” in potential d-luciferin interacting sites (DLS) and a high amino acid identity to the luciferase of American firefly Photinus pyralis (Ppy-luc1) (> 47%). Luminous clades are highlighted in green.

Genetic basis of light on/off and its pattern. To explore genetic basis underlying light on/off and its pattern (i.e. glow, flash), we combined comparative genomics with transcriptomic and proteomic data of luminous organs of glow (L. yunnana) and flash (A. terminalis) taxa to investigate the gene families (Fig. 7a; Supplementary Note 9, Tables S54–S58, Figs. S68–S89, Data S32) related to the previously reported flash control model and cell calcium signaling pathway because calcium ions were involved in the intense, long lasting scintillation in Photuris fireflies. Our results indicate that two positively selected genes (sarcoplasmic/endo-
plasmic reticulum calcium-transporting ATPase and calreticulin) of the calcium signaling pathway in the ancestor of luminous beetles (Fig. 7b,c) and voltage-dependent anion channel (VDAC) have a strongly expression at transcriptomic and proteomic levels (Fig. 7d–g), especially for VDAC that showing a higher expression in flash firefly *A. terminalis* than in glow firefly *L. yunnana* (Fig. 7h). Sarcoplasmic/endoplasmic reticulum calcium-transporting ATPase, calreticulin and voltage-dependent anion channel play an important role in calcium signal between mitochondria and reticulum \[^{50–52}\]. Additionally, our transcriptomic data also show that most of other genes in the previously reported flash model \[^{9,10}\] (i.e. octopamine receptors, one of α subunit genes of Guanine nucleotide-binding (G) proteins (Gs), adenylyl cyclases, cAMP-dependent protein kinase) generally have a high expression (Supplementary Fig. S86, Data S33). Thus, it raised doubt on whether nitric oxide synthase plays a key role as synthase in flash fireflies, while M/L/V/R/A is in the same position of glow beetles and non-luminous taxa non-luminous insects shows that a specific amino acid site (Q) exists in the oxygenase domain of nitric oxide synthase \[^{30}\]. Those genes with abundance at proteomic level were shown bold in (b, d). (f–i) The ventral view of CRISPR/Cas9-induced morphological mutants (g, i: no luminous organs and the regions correspondent to luminous organs of wild type highlighted in red in (g)) of Abdominal-B in *A. terminalis* and their WT (f, h: having luminous organs labelled in green circled area in (f) and shown blue spots in (h)). The numbers show the abdominal 7 segments (A7)–A10 and abnormal segments were shown in red number; f–g: observed using the microscope SMZ1000 (Nikon, Japan) and photos were taken using Canon 700D; h–i: three-dimension reconstruction of A7–A10. (j) Expression of candidate genes between mutants and wide types. The pink stars and blue circles show the significant (|log2(FC)|> 2, *P* < 0.01) up-regulated genes and down-regulated genes in *A. terminalis* mutants, respectively. (k) The stereoisomeric conversion of l-luciferin to d-luciferin in vitro by luciferase and acyl-CoA thioesterase (ACOT). The three HPLC (high-performance liquid chromatography) chromatograms represent standard l-luciferin (up), standard d-luciferin (middle) and stereoisomeric inversion of l-luciferin to d-luciferin in vitro by luciferase and acyl-CoA thioesterase (ACOT) (down).

**Discussion**

Our comprehensive investigation by integrating multilevel data provides multiple insights into the origin of luciferin and bioluminescence as well as firefly phylogeny. Our results clarify that the origin and evolution of luciferase genes play a leading role in the origin of luciferin and thus bioluminescence. Our experimental results demonstrate that one of acyl-CoA thioesterases can efficiently convert l-luciferin to d-luciferin (the substrate for bioluminescence reaction). Our phylogenomic analyses reveal a closer phylogenetic position of *Lamprigera* to Luciolinae, and thus *L. yunnana* should be a member of Luciolinea rather than a species in Lampyridae. However, due to no available reference genomes of representatives from other families of Elateroidea (e.g., Rhagophthal- midae, Phengodidae, Cantharidae, Lycidae) and from non-luminous species in Elateridae now, a more expanded comparative genomics to include these related taxa will still be needed to clarify more details on the evolution of luciferase and other genes in luciferin biosynthesis and their contribution to bioluminescence origin. In addition, due to such natures of fireflies as long life cycle, difficulty to rear in large scale in the laboratory etc., we were only able to get mutants of *Abd-B* knock-out by gene editing, while all other functional verification efforts on more

---

\[^{11}\]
Figure 6. The evolution of acyl-CoA thioesterases (ACOTs) in beetles and other organisms. (a) The phylogenetic tree, motifs and domains. The tree was divided into three clusters, I, II and III. In cluster I, ① shows the clade of single copy syntrophic orthologs (see b) (also as candidates in luciferin biosynthesis) in luminous beetles, ② shows the specific clade in insect ACOTs and ③ shows the clade of mammals ACOTs locating in mitochondria. The highlighted three ACOTs (AteACOT1, AteACOT4 and AteACOT9) were selected as candidates to verify their function in converting L-luciferin to D-luciferin in vitro experiments. The red circles show those genes with the C-terminal peroxisomal targeting signal 1 (PTS1). The probability of motifs was shown using P value. The domains were identified using Pfam database. 4HBT and 4HBT_3 domains belong to CL0050 Hotdog superfamily. E.coliACOTII shows fatty acyl-CoA thioesterase (ACOT) II (TESB) from *Escherichia coli*. b shows the syntenics relationships of ACOT gene highlighted in pink and its flanking genes surrounding up-200 kb and down-200 kb genomic regions among six luminous beetles. Lyu: Lamprigera yunnana; Ate: Abscondita terminalis; Ala: Aquatica lateralis; Ppy: Photinus pyralis; Ppe: Pyrocoelia pectoralis; Ilu: Ignater luminosus; Apl: Agrilus planipennis; Ota: Onthophagus taurus; Agl: Anoplophora glabripennis; Dpo: Dendroctonus ponderosae; Tca: Tribolium castaneum; Dme: Drosophila melanogaster; Mus: Mus musculus; Homo: Homo sapiens.

genes, especially related to luciferin biosynthesis were proved to be tremendously difficult. More exploration on lab-rearing and gene editing of fireflies are needed to collect accurate data to testify the biosynthesis pathway of luciferin which we propose here in the future. This study has laid a pivotal foundation for future studies on all luminescent insect taxa together with efficient functional assays to completely reveal all mysteries underlying the fascinating phenomenon of firefly and all luminescent insects’ bioluminescence ever since Aristotle and Darwin.

Methods

Firefly collection, breeding and sample treatment. Adults and larvae of Lamprigera yunnana were collected in Kunming City, Yunnan, China from 2014 to 2016. Adults of Abscondita terminalis were collected in Menglin, Xishuanbana, Yunnan, China from 2015 to 2018. Live fireflies were brought in plastic containers back to the lab for observing their biological and morphological traits, breeding and storing samples frozen in liquid nitrogen until used. The last instar larva of *L. yunnana* at lower instars were got by feeding lower instar larva collected from wild with snails from the same habitat in covered plastic boxes (16 × 10 × 5 cm). Female and male pair of *A. terminalis* were transferred into covered plastic boxes (16 × 10 × 5 cm) padded with a wet paper napkin in an incubator at 25–27 °C and 70% relative humidity. The newly laid eggs were collected for continuous breeding at the same condition or for gene editing experiment. Hatched larvae were fed with chopped *Tenebrio molitor* larva to get different instar larvae and pupae. Adults collected in the wild were carefully dissected to get whole body but excluding wings (in case of pterygote stages), or only separated luminous organs, and frozen at −80 °C until used.

The whole bodies of single *L. yunnana* female adult and single *A. terminalis* female were used for genome survey using Illumina sequencing technology. The whole bodies of another single *L. yunnana* female adult and another 15 *A. terminalis* female adults were used for de novo sequencing using single-molecular real-time (SMRT) technology in the PacBio platform. The whole body of single individual of both species at different developmental stages (*L. yunnana*: larvae, male adult and female adult; *A. terminalis*: larvae, male pupae, female pupae, male adult, and female adult), and luminous organs of both sex for both species (*L. yunnana*: mixed 3 individuals; *A. terminalis*: mixed 10 individuals) were used for transcriptomic sequencing. The luminous organs of both sex for both species (*L. yunnana*: mixed 6 individuals; *A. terminalis*: mixed 70 female individuals, mixed 40 male individuals) were used for proteomic sequencing.

Genome sequencing and assembly. Genomic DNA for genome survey (Illumina sequencing) was extracted from the whole body of single female adult for both *L. yunnana* and *A. terminalis* using a Gentra Puregene Blood Kit (Qiagen, Germany) following manual instructions. The libraries of the 350 bp short insert were sequenced on the Illumina HiSeq4000 to obtain pair-end reads, which was used to estimate genomic characteristics based on k-mer frequency distribution (Supplementary Note 2) using a similar method as described previously24, and also used to polish assembled genomes based on only PacBio reads. For de novo sequencing (PacBio Sequencing), high-molecular-weight genomic DNA was extracted from the whole body of single *L. yunnana* female adult and 15 *A. terminalis* female adults with Sodium Dodecyl Sulfonate method, and the 20 kb libraries (four for *L. yunnana* and one for *A. terminalis*) were constructed and sequenced with a PacBio RS II platform (Pacific Biosciences, USA) using the P6 polymerase/C4 chemistry combination. A long noisy reads assembler, wtdbg15 (the source code is available on GitHub: https://github.com/ruanjue/wtdbg) was selected to assemble the genomes of two species as the followings (Supplementary Note 2). Firstly, using wtdbg, we performed the primary genome assembly of both species, followed by the first round of polishing using the wtdbg-cns program to produce the polished contigs and then the second round of polishing by combining minimap with wtdbg-cns to obtain the preliminary contigs. Secondly, the Quiver26 within SMRT Analysis v2.3.0 was used to polish base calling of preliminary contigs to improve the site-specific consensus accuracy of the assembly. Finally, we applied for the program Pilon27 with the “fix-all” mode to implement two consecutive rounds of polishing using Illumina short reads (Supplementary Table S1) to achieve the final assembly. The Illumina short reads, assembled transcripts and Benchmarking Universal Single-Copy Orthologs (BUSCO) were used to evaluate the completeness of assemblies (Supplementary Note 2).
Genome annotation. Repetitive elements (transposable elements (TEs) and tandem repeats) were annotated using a combined strategy of de novo-based prediction, homology-based approach and Tandem Repeat Finder (TRF) in *L. yunnana* and *A. terminalis* (Supplementary Note 3). Gene model identification was conducted by a combination of de novo prediction, homology-based prediction and transcriptome-based prediction methods, and gene functional annotations were performed using BLASTP (E-value < 1e−5) against SwissProt, TrEMBL and NCBI non-redundant protein Database (NR) (Supplementary Note 3). Genes were extracted based on the best BLAST hit along with their protein functional annotation. Structural protein domains and motifs were searched against SMART, ProDom, Pfam, PRINTS, PROSITE and PANTHER databases using InterProScan v5.25. Gene Ontology (GO) terms for genes were obtained from the corresponding InterPro entry. The metabolic pathways in which the genes might be involved were assigned by BLAST against the KEGG model following the author's recommendation.

Phylogenetic analysis and genome evolution. Gene families and single-copy orthologs were constructed using the OrthoMCL-based on all-to-all BLASTP (E-value ≤ 1e−5) alignments. The phylogenetic tree of single-copy genes was constructed using RAxML v8.0 under the GTR + gamma model, and their divergence times were estimated using the PAML v4.8 program. 15 beetles (6 luminous taxa in Elateroidea (5 fireflies, 1 luminous click beetles) and other 9 non-luminous beetles) with genomes available were included in the phylogenetic analysis with *Drosophila melanogaster* (Dme) as an outgroup (Supplementary Note 4). Besides, the mitogenomic phylogenetic tree of 39 Elateroidea species including five of above mentioned was also inferred with *T. castaneum* as an outgroup (Supplementary Note 4). To perform extensively comparison of genomes, 20 insect species (11 beetle species after removing 4 non-Elateroidea taxa of above mentioned 15 species because of their poor assembly and annotation; Lepidoptera: 3; Diptera: 1; Hymenoptera: 2; Hemiptera: 1; Phthiraptera: 1; Isoptera: 1) plus one outgroup (Crustacean, Cladocera: *Daphnia pulex*) were included in the phylogenetic tree of luminous organs of adult females and males of *L. yunnana* (*Lyu*) and *A. terminalis* (*Ate*) and *L. yunnana* and *A. terminalis*.

Transcriptome sequencing, proteome sequencing and analysis. Total RNA was extracted using the guanidinium thiocyanate–phenol–chloroform extraction method (Trizol, Invitrogen) according to the manufacturer's protocol. RNA sequencing libraries (350 bp insert size) were generated using Illumina mRNA-Seq Prep Kit and sequenced using Illumina HiSeq4000 sequencer with read length of PE150. Two methods, i.e., de novo assembly of clean reads using Bridger with the default setting (k-mer size of 25) and mapping them back to the assembled genomes using TopHat, were carried out for transcriptome assembly. The correlation of global expression (reads count) among samples was analyzed using cor function with spearman method from R program. The fragments per kilobase of exon per million fragments mapped (FPKM) values were calculated using Cufflinks software package and used to measure gene expression. The genes were remained with the total...
FPKM > 0 from all samples as expressed genes (EGs). The high expression genes (HEGs) were determined by choosing the first 5% EGs ranked from high to low based on the expression in luminous organs. The differentially expressed genes (DEGs) between luminous organs were analyzed using the EdgeR program. The genes with the absolute value of logFC ≥ 4 and false discovery rate (FDR) ≤ 0.01 were identified as DEGs of interspecific luminous organs due to longer species divergence time.

Total protein from each luminous organ sample was prepared, and 10 μg of proteins from each sample were used for tryptic digestion. The peptide samples were labelled using iTRAQ kits (Applied Biosystems, Foster City, CA) and analyzed using TripleTOF 5600+ mass spectrometer coupled with the Eksigent nanoLC System (SCIEX, USA). Protein identification and quantification were performed using ProteinPilot 4.5 software. The correlation of quantitative results was evaluated using Pearson algorithm. The high abundance proteins (HAPs) were determined with the first 5% proteins ranked from high to low based on the abundance in luminous organs. The different abundance proteins (DAPs) were defined with a fold-change (FC) ≥ 2 or ≤ 0.5 and a P value ≤ 0.05 (t-test of all comparison groups). The R package (https://www.r-project.org/) was used for statistical expression data and visualization.

**Bioluminescence gene families and pathways.** To explore the origin of bioluminescence in fireflies, we summarized and expanded the pathway of metabolism of luciferin, the emitter of light, which only exist in luminous insects, and investigated the candidate genes in the expanded luciferin metabolism (especially its biosynthesis) in the genomes of luminous beetles and non-luminous beetles (Supplementary Note 7). To explore the possible molecular mechanism of flash on/off and their difference between taxa as well as the possible contribution of calcium to flash control, we thoroughly investigate the candidate genes in the pathways of cAMP/PKA-Ca/Calmodulin signaling cascade and related calcium signaling (Supplementary Note 9). The expression of these candidate genes identified in luminous organs was analyzed using R software based on the FPKM values calculated using Cufflinks. The phylogenetic trees were constructed using RAxML with maximum likelihood method.

**Microsynteny analysis.** To explore the homologous systemic blocks in luciferin biosynthesis (major candidate genes), we performed the identification of genome-wide syntenic and collinear blocks across the six luminous species (Lampyridae: *L. yunnana*, *A. terminalis*, *A. lateralis*, *P. pyralis* and *P. pectoralis*; Elateridae: *I. luminosus*). First, the database of protein similarity was obtained based on all-to-all Blastp (-evalue 1e-10, -num_alignments 20) of the translated protein sequences from six luminous beetles. Second, we used the Multiple Collinearity Scan (Mcscan) (Mcscan toolkit version 1.1, 2016) with more than 3 homologous gene pairs per block to identify conserved collinear blocks, generating a syntenic or collinear block database across all of six species. To perform synteny analysis, we searched and located the target genes along the collinear blocks with the flanking genes surrounding up-200 kb and down-100/200 kb genomic regions as well as the counterparts from different genome. In addition, for the target genes and their flanking genes absent in collinear blocks, we manually scanned the protein similarity database and regarded the gene pairs from different species with more than 50% identity and 80% coverage as the synteny. The syntenic relationships of genes in luciferin biosynthesis and their flanking genes between six luminous species were visualized using Mcscan (https://github.com/tangalbio/jcvi/wiki/Mcscan-%28Pythonversion%29#dependencies).

**Functional verification of genes in luciferin deracemization in vitro.** The coding sequences of the firefly luciferase (LUC), alpha-methyl-acyl-CoA-racemase (AMACR) and acyl-CoA thioesterases (ACOT) of *A. terminalis* were synthesized and separately constructed into pET-28a vector (Takara, Japan). ACOTs were further subcloned into pCold-TF vector (Takara, Japan) because they failed to be well expressed in pET-28a vector. The LUC, AMACR and ACOT were expressed in *E. coli* BL21(DE3) at 15 °C. Then, the proteins were purified using nicke in trilitriacetic acid (Ni²⁺-NTA) column (Qiagen, Germany) and used for the following experiments.

The in vitro deracemization reaction mixture (200 μL) contained 0.1 mM l-luciferin, 8 mM MgSO₄, 3 mM ATP, 0.5 mM CoASH, and each (1 μg) of enzymes (LUC, AMACR, ACOT) in 100 mM Tris–HCl (pH = 8.0). The reaction time was 45 min at 30 °C. The chirality of luciferin was monitored by high-performance liquid chromatography (HPLC) system (Alliance HPLC System with 2695 Separation Module, 2475 Multi λ Fluorescence Detector, Waters) using a chiral fused silica column (Chiralcel OD-RH, 4.6 × 150 mm; Daicel Chemical Industry, Tokyo, Japan). d-luciferin and l-luciferin were detected with a fluorescence detector (excitation λ = 330 nm, emission λ = 530 nm) (Supplementary Note 7). Relative light units (RLU) of the reaction mixture were measured using a Luminescer Octa AB-2270 (ATTO, Tokyo, Japan). During 20 s, the integrated activity was described using Pearson algorithm. The high abundance proteins (HAPs) were determined with the first 5% proteins ranked from high to low based on the abundance in luminous organs. The different abundance proteins (DAPs) were defined with a fold-change (FC) ≥ 2 or ≤ 0.5 and a P value ≤ 0.05 (t-test of all comparison groups). The R package (https://www.r-project.org/) was used for statistical expression data and visualization.

**Gene editing in firefly using CRISPR/Cas9 system.** Considering that Hox gene *Abd-B* is related to luminous organ development, we selected the homeobox region of the *Abd-B* gene to perform CRISPR/Cas9 gene editing in *A. terminalis*. Target site selection and sgRNA preparation mainly follow the methodology in our previous studies (Supplementary Note 8.1, Supplementary Fig S61). Recombinant Cas9 protein (PNA Bio Inc, CA, USA) was used.

During the day, females and males collected from the wild were reared in covered plastic boxes (16 × 10 × 5 cm) padded with a wet paper napkin in an incubator at 25–27 °C and 70% relative humidity. After 0 o’clock in the evening, females were moved into a new covered plastic box (16 × 10 × 5 cm) containing mosses with sufficient humidity and completely dark for oviposition of 7–8 h. In the next morning, fresh *A. terminalis* eggs were collected from the soaked moss by repeated pipetting with a pipette, and then pipetted and arranged on a
microscope slide (25.4 × 76.2 × 1–1.2 mm). We injected ~ 2 nl of the mixture of sgRNAs and Cas9 protein (PNA Bio, CA, USA) into each egg under a dissecting microscope (SMZ 800, Nikon, Japan) using a TransferMan NK2 equipped with a TwinTip Holder and Femtojet microinjection system (Eppendorf, Germany) at 16–18 °C. Injection needles were made of glass capillary (100 × 1 × 0.6 mm, BJ-40, Zheng-Tian-Yi, Beijing, China) by Narishige PN 30 (Japan) under the parameters: Heater: 80 °C, Magnet Sub: 40 °C, Magnet Main: 50 °C. Optimaly, egg injection should be undertaken as early as possible, e.g., at the “one nucleus” stage. Based on the hatching time of about two weeks for A. terminalis eggs, all experiment steps, from egg laying to injection, should be finished within 10 h after egg laying (AEL). After injection, the eggs on the slides were carefully washed into covered plastic boxes (16 × 10 × 5 cm) padded with a wet paper napkin, and then placed in an incubator at 25–27 °C and 70% relative humidity. The hatched larvae (generation 0, G0) were carefully moved to clean plastic box padded with wet paper napkin with chopped Tenebrio molitor larvae as food. The phenotype of G0 larva was carefully checked especially in the abdomen using microscope SMZ 645 and SMZ 18 (Nikon, Japan). The morphologically abnormal individuals and the un-injected wild type individuals were photographed using an AMZ 100 system with a digital camera (Nikon, Japan).

Genotyping was carried out for single or mixed injected larvae (Supplementary Note 8). Genomic DNA extraction from the whole body of single or mixed larvae and subsequent PCR were carried out using TransDirect Animal Tissue PCR kit (TransGen, Beijing, China) following the manufacturer’s instructions. The primer pairs were the same as mentioned above, i.e., LT07795_ex2-F1/R1 for second exon and LT07795_ex3-F3/R3 third exon (Supplementary Fig. 64). The PCR products of the target sites were cloned into pMD-19 T (Takara, Japan), and 10 clones for each sample were selected for Sanger sequencing. Successfully sequenced data were aligned and analyzed using Lasergene SeqMan Pro software (version 7.1) (DNASTAR).

Mixed 14 mutants and mixed 14 wild type larvae were used for total RNA extraction, respectivey, and RNA sequencing according to methodology above description (Supplementary Note 8.3). The clean reads were first mapped to the de novo assembled genome of A. terminalis using Tophat49, and then used to calculate FPKM values and analyze the differentially expressed genes (DEGs) between mutant larva and wild type using Cufflinks software package50.

In addition, we also tried CRISPR/Cas9 gene editing for some genes in proposed biosynthesis pathway, i.e. luciferase. However, due to such natures of fireflies as long life circle, difficulty to rear in large scale in the laboratory etc., we still couldn’t obtain enough injected larvae for phenotyping at current time. More technologies about firefly raise in large scale still need to develop and more improved gene editing experiments need to be performed for further testing the function of these genes.

Data availability
The genome assemblies and sequence data, RNA-seq data for Lamprigera yunnana and Absconidia terminalis were deposited at NCBI under BioProject accession number PRJNA556754 and PRJNA556938, respectively. The quantitative proteome data for L. yunnana and A. terminalis were deposited at iProX under Project accession number IPX0001742000 (PXD015226) and IPX0001743000 (PXD015227), respectively.

Received: 13 March 2020; Accepted: 9 September 2020
Published online: 28 September 2020

References
1. Harvey, E. N. *Bioluminescence* (Academic Press, New York, 1952).
2. Darwin, C. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (John Murray, London, 1859).
3. Buck, J. B. The anatomy and physiology of the light organ in fireflies. *Ann. NY Acad. Sci.* 49, 397–483 (1948).
4. Nakatsu, T. et al. Structural basis for the spectral difference in luciferase bioluminescence. *Nature* 440, 372–376 (2006).
5. Viviani, V. R. The origin, diversity, and structure function relationships of insect luciferases. *Cell. Mol. Life Sci.* 59, 1833–1850 (2002).
6. Oba, Y., Shintan, T., Nakamura, T., Ojika, M. & Inouye, S. Determination of the luciferin contents in luminous and non-luminous beetles. *Biosci. Biotechnol. Biochem.* 72, 1384–1387 (2008).
7. Branhm, M. A. & Wenzel, J. W. The origin of photic behavior and the evolution of sexual communication in fireflies (Coleoptera: Lampyridae). *Cladistics* 19, 1–22 (2003).
8. Branhm, M. A. & Greenfield, M. D. Flashing males win mate success. *Nature* 381, 745–746 (1996).
9. Aprille, J. R., Lagace, C. J., Modica-Napolitano, J. & Trimmer, B. A. Role of nitric oxide and mitochondria in control of firefly flashing. *Mol. Phylogenet. Evol.* 140, 106600 (2019).
10. Stern, D. L. The genetic causes of convergent evolution. *Nat. Rev. Genet.* 14, 751–764 (2013).
11. Kundrata, R., Bocakova, M. & Bocak, L. The comprehensive phylogeny of the superfamily Elateroidea (Coleoptera: Elateriformia). *ZooKeys* 864, 19 (2019).
12. Zhang, S. Q. et al. Evolutionary history of Coleoptera revealed by extensive sampling of genes and species. *Nat. Commun.* 9, 1 (2018).
13. Chen, X. et al. Phylogenetic analysis provides insights into the evolution of Asian fireflies and adult bioluminescence. *Mol. Phylogenet. Evol.* 140, 106600 (2019).
14. Liu, G. C. et al. Nitric oxide and the control of firefly flashing. *Science* 292, 2486–2488 (2001).
15. Fu, X. et al. Long-read sequence assembly of the firefly Pyrocoelia pectoralis genome. *Gigascience* 6, 1–7 (2017).
16. Fallon, T. R. et al. Firefly genomes illuminate parallel origins of bioluminescence in beetles. *Elife* 7, e36495 (2018).
17. Liu, G. C. et al. Genome size of 14 species of insects (Insecta, Coleoptera, Lampyridae). *Zool. Res.* 38, 449–458 (2017).
18. Kundrata, R., Bocakova, M. & Bocak, L. The comprehensive phylogeny of the superfamily Elateroidea (Coleoptera: Elateriformia). *Mol. Phylogenet. Evol.* 76, 162–171 (2014).
19. McKenna, D. D. et al. The beetle tree of life reveals that Coleoptera survived end-Permain mass extinction to diversify during the Cretaceous terrestrial revolution. *Syst. Entomol.* 40, 835–880 (2015).
20. Bi, W. X., He, J. W., Chen, C. C., Kundrata, R. & Li, X. Y. Sinopryrophorinae, a new subfamily of Elateridae (Coleoptera, Elateroidae), with the first record of a luminous click-beetle in Asia and the evidence for multiple origins of bioluminescence in Elateridae. *Zookeys* 864, 19 (2019).
21. Zhang, S. Q. et al. Evolutionary history of Coleoptera revealed by extensive sampling of genes and species. *Nat. Commun.* 9, 1 (2018).
22. Chen, X. et al. Phylogenetic analysis provides insights into the evolution of Asian fireflies and adult bioluminescence. *Mol. Phylogenet. Evol.* 140, 106600 (2019).
23. Stern, D. L. The genetic causes of convergent evolution. *Nat. Rev. Genet.* 14, 751–764 (2013).
21. Jeng, M. L., Lai, J., Yang, P. S. & Sato, M. Notes on the taxonomy of Lamprigera yunnana (Fairmaire) and the genus Lamprigera Motschulsky (Coleoptera, Lampyridae). **Ipsa. J. Syst. Entomol.** **6**, 313–319 (2000).

22. Martin, G. J., Branham, M. A., Whiting, M. F. & Bybee, S. M. Total evidence phylogeny and the evolution of adult bioluminescence in fireflies (Coleoptera: Lampyridae). **Mol. Phylogenet. Evol.** **107**, 564–575 (2017).

23. Kazantsev, S. V. Protocelciula albertalli gen.n., sp.n., a new Luciliinae firefly (Insecta: Coleoptera: Lampyridae) from Burmite amber. **Russ. Entomol. J.** **24**, 281–283 (2015).

24. Khurana, P., Gokhale, R. S. & Mohanty, D. Genome scale prediction of substrate specificity for acyl adenylate superfamily of enzymes based on active site residue profiles. **BMC Bioinform.** **11**, 57 (2010).

25. Day, J. C., Goodall, T. L. & Bailey, M. J. The evolution of the adenylate-forming protein family in beetles: multiple luciferase gene paralogues in fireflies and glow-worms. **Mol. Phylogenet. Evol.** **50**, 93–101 (2009).

26. Conti, E., Franks, N. P. & Brick, P. Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. **Structure** **4**, 287–298 (1996).

27. Ohta, Y., Ida, K. & Inouye, S. Functional conversion of fatty acyl-CoA synthetase to firefly luciferase by site-directed mutagenesis: a key substitution responsible for luminescence activity. **Festschr.** **583**, 2004–2008 (2009).

28. Yongsangnak, W., Chunamanpun, P. & Siriboonlert, A. Transcriptome analysis reveals candidate genes involved in luciferin metabolism in *Luciola aquatilis* (Coleoptera: Lampyridae). **PeerJ** **4**, e2534 (2016).

29. Okada, K., Ito, H. & Goto, T. Biosynthesis of firefly luciferin. Probable formation of benzothiazole from para-benzoquinone and cysteine. **J. Chem. Soc. Chem. Commun.** **1**, 32 (1976).

30. Kanie, S., Nishikawa, T., Ojika, M. & Oba, Y. One-pot non-enzymatic formation of firefly luciferin in a neutral buffer from p-benzoquinone and cysteine. **Sci. Rep.** **6**, 24794 (2016).

31. Streheier, B. L., Press, G. D. & Raychaudhuri, A. Histochemistry of the lantern of the firefly *Photinus pyralis* (Coleoptera: lampyridae). **Am. Entomol. Soc. Ann.** **60**, 81–91 (1967).

32. Oba, Y., Yoshida, N., Kanie, S. & Inouye, S. Biosynthesis of firefly luciferin in adult lantern: decarboxylation of L-cysteine is a key step for benzothiazole ring formation in firefly luciferin synthesis. **PLoS ONE** **9**, e95063 (2014).

33. Taylor, A. M., Kammath, V. & Bleakley, A. Tyrosinase, could it be a missing link in ochronosis in alkaptonuria?. **PLoS ONE** **9**, 564–575 (2014).

34. Niwa, K., Nakamura, M. & Ohmiya, Y. Stereoisomeric bio-inversion key to biosynthesis of firefly d-luciferin. **Sci. Rep.** **5**, 1269–1271 (2016).

35. Zannoni, V. G., Lomtevas, N. & Goldfinger, S. Oxidation of homogentisic acid to ochronotic pigment in connective tissue. **Biochemistry** **45**, 117–128 (2006).

36. Merilainen, G., Poikela, V., Kursula, P. & Wierenga, R. K. The thiolase reaction mechanism: the importance of Asn316 and His348 for stabilizing the enolate intermediate of the Claisen condensation. **Biochemistry** **48**, 11011–11025 (2009).

37. Seedorf, U., Brysch, P., Engel, T., Schrage, K. & Assmann, G. Sterol carrier protein-X is peroxisomal 3-oxoacyl coenzyme A thiolase for stabilizing the enolate intermediate of the Claisen condensation. **Biochemistry** **48**, 1397–1410 (2012).

38. Brocker, C., Carpenter, C., Nebert, D. W. & Vasiliou, V. Evolutionary divergence and functions of the human acyl-CoA thioesterase family (ACOT) family. **Hum. Genom.** **4**, 411–420 (2010).

39. Cantu, D. C., Ardevol, A., Rovira, C. & Reilly, P. J. Molecular mechanism of a hotdog-fold acyl-CoA thioesterase. **Chemistry** **20**, 9045–9051 (2014).

40. Subramani, S. Components involved in peroxisome import, biogenesis, proliferation, turnover, and movement. **Physiol. Rev.** **78**, 171–188 (1998).

41. Fallon, T. R., Li, F. S., Vicent, M. A. & Weng, J. K. Sulfoluciferin is biosynthesized by a specialized luciferin sulfortransferase in fires. **Biochemistry** **55**, 3341–3344 (2016).

42. Hunt, M. C., Siponen, M. I. & Alexson, S. E. The emerging role of acyl-CoA thioesters and acyltransferases in regulating peroxisomal lipid metabolism. **Biochim. Biophys. Acta** **1822**, 1397–1410 (2012).

43. Brocke, C., Carpenter, C., Nebert, D. W. & Vasiliiou, V. Evolutionary divergence and functions of the human acyl-CoA thioesterase family (ACOT) family. **Biochim. Biophys. Acta** **1822**, 1374–1386 (2012).

44. Hunt, M. C., Tillander, V. & Alexson, S. E. H. Regulation of peroxisomal lipid metabolism: the role of acyl-CoA and coenzyme A metabolizing enzymes. **Biochimica** **95**, 44–55 (2014).

45. Carlson, A. D. Is the firefly flash regulated by calcium??. **Integr. Comp. Biol.** **44**, 220–224 (2004).

46. Corbett, E. F. et al. Cal**-**** regulation of interactions between endoplasmic reticulum chaperones. **J. Biol. Chem.** **274**, 6203–6211 (1999).

47. Michalak, M., Groenendyk, I., Szabo, E., Gold, L. I. & Ops, M. Calreticulin, a multi-process calcium-buffering chaperone of the endoplasmic reticulum. **Biochem. J.** **417**, 651–666 (2009).

48. Kuhlbrandt, W. Biology, structure and mechanism of P-type ATPases. **Nat. Rev. Mol. Cell. Biol.** **5**, 282–295 (2004).

49. Ohtsuki, H., Yokoyama, J., Ohba, N., Ohmiya, Y. & Kawata, M. Expression of the nos gene and firefly flashing: a test of the nitric-oxide-mediated flash control model. **J. Insect Sci.** **14**, 56 (2014).

50. Li, Y. & et al. Outbred genome sequencing and CRISPR/Cas9 gene editing in butterflies. **Nat. Commun.** **6**, 1 (2015).

51. Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. **Nat. Methods** **17**, 155–158 (2020).

52. Chm, G. S. et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. **Nat. Methods** **10**, 563–569 (2013).

53. Walker, B. J. et al. Plsm: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. **PLoS ONE** **9**, e121963 (2014).

54. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. **Nucleic Acids Res.** **27**, 49–54 (1999).

55. Apweiler, R. et al. Ongoing and future developments at the universal protein resource. **Nucleic Acids Res.** **39**, D214–D219 (2011).

56. Jones, P. et al. InterProScan: 5: genome-scale protein families, domains and functional classification through signature alignment. **Bioinformatics** **24**, 52–62 (2008).

57. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. **Nucleic Acids Res.** **28**, 27–30 (2000).

58. Li, L., Stoeckert, C. J. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. **Genome Res.** **13**, 2178–2189 (2003).

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

60. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. **Mol. Biol. Evol.** **24**, 1586–1591 (2007).

61. De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFe: a computational tool for the study of gene family evolution. **Bioinformatics** **22**, 1269–1271 (2006).
Author contributions
X.L., W.W. conceived the study, designed scientific objectives, led the project and manuscript preparation. R.Z., X.L., W.W. conceived the study, designed scientific objectives, led the project and manuscript preparation. R.Z., X.L., W.W., J.H., S.G., Z.D., G.L., X.C., W.X., K.D., F.H., R.Z. performed CRISPR/Cas9 gene editing. X.L., R.Z., W.W., J.H., S.G., Z.D., G.L., X.C., W.X., K.D., F.H., R.Z. performed the in vitro cell experiments. J.H., Z.D., G.L., X.L., R.Z. performed CRISPR/Cas9 gene editing. X.L., R.Z., W.W., J.H., S.G., Z.D., X.Z. wrote the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-72900-z.

Correspondence and requests for materials should be addressed to S.G., W.W. or X.L.

Reprints and permissions information is available at www.nature.com/reprints.

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

© The Author(s) 2020