Gene 850 (2023) 146917

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

# Genome analysis of *Phrixothrix hirtus* (Phengodidae) railroad worm shows the expansion of odorant-binding gene families and positive selection on morphogenesis and sex determination genes

Danilo Trabuco Amaral [a], Yasuo Mitani [b], Isabel Aparecida Silva Bonatelli [c], Ricardo Cerri [d], Yoshihiro Ohmiya [e,f], Vadim Ravara Viviani [g,h,*]

[a] *Programa de Pós-Graduação em Biotecnociência, Centro de Ciências Naturais e Humanas. Universidade Federal do ABC (UFABC), Santo André, Brazil*
[b] *Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Sapporo, Japan*
[c] *Department of Ecology and Evolutionary Biology, Universidade Federal de São Paulo (UNIFESP), Diadema, São Paulo, Brazil*
[d] *Department of Computer Science, Universidade Federal de São Carlos (UFSCar), São Carlos, Brazil*
[e] *Biomedical Research Institute, AIST, Ikeda-Osaka, Japan*
[f] *Osaka Institute of Technology, OIT, Osaka, Japan*
[g] *Graduate Program of Evolutive Genetics and Molecular Biology, Federal University of São Carlos (UFSCar), São Carlos, Brazil*
[h] *Graduate Program of Biotechnology and Environmental Monitoring, Federal University of São Carlos (UFSCar), Sorocaba, Brazil*

ARTICLE INFO

*Keywords:*
Bioluminescence
Elateriform
Genomic DNA
Morphogenesis
*Odorant-binding protein*
Phengodidae

ABSTRACT

Among bioluminescent beetles of the Elateroidea superfamily, Phengodidae is the third largest family, with 244 bioluminescent species distributed only in the Americas, but is still the least studied from the phylogenetic and evolutionary points of view. The railroad worm *Phrixothrix hirtus* is an essential biological model and symbolic species due to its bicolor bioluminescence, being the only organism that produces true red light among bioluminescent terrestrial species. Here, we performed partial genome assembly of *P. hirtus,* combining short and long reads generated with Illumina sequencing, providing the first source of genomic information and a framework for comparative analyses of the bioluminescent system in Elateroidea. This is the largest genome described in the Elateroidea superfamily, with an estimated size of ~3.4 Gb, displaying 32 % GC content, and 67 % transposable elements. Comparative genomic analyses showed a positive selection of genes and gene family expansion events of growth and morphogenesis gene products, which could be associated with the atypical anatomical development and morphogenesis found in paedomorphic females and underdeveloped males. We also observed gene family expansion among distinct *odorant-binding receptors*, which could be associated with the pheromone communication system typical of these beetles, and *retrotransposable elements*. Common genes putatively regulating bioluminescence production and control, including two luciferase genes corresponding to lateral lanterns green-emitting and head lanterns red-emitting luciferases with 7 exons and 6 introns, and genes potentially involved in luciferin biosynthesis were found, indicating that there are no clear differences about the presence or absence of gene families associated with bioluminescence in Elateroidea.

## 1. Introduction

Among the three main families of bioluminescent beetles, the less studied and most enigmatic one is Phengodidae, which contains the so-called railroad worms. This family has 35 genera and about 200 species distributed in the Americas in three tribes, Pennicilloporini, Phengodini, and Mastinocerini; the latter being distributed mainly in the Neotropical region (Wittmer, 1976; Viviani and Bechara, 1997; O'Keefe et al., 2002; Viviani, 2002). One of the differential and unique characteristics of these atypical beetles is the occurrence of paedomorphy in adult

*Abbreviations:* ACT, *acyl-CoA thioesterase*; Gb, giga base paris; GO, gene ontology; LRE, *luciferin-regenerating enzyme*; mtDNA, mitochondrial genome; Mya, millions year; OBP, *odorant-binding protein*; PSG, positive selection genes; rnaL, ribosomal RNA large unit; rnaS, ribosomal RNA small unit; SOD, *superoxide dismutase*; TE, transposable element; tRNA, transport RNA.

* Corresponding author at: Rodovia João Leme dos Santos, Km 110, SP264. ZIP 18052-780, Sorocaba, Brazil.
*E-mail address:* viviani@ufscar.br (V.R. Viviani).

https://doi.org/10.1016/j.gene.2022.146917
Received 8 August 2022; Received in revised form 14 September 2022; Accepted 21 September 2022
Available online 27 September 2022

females, underdeveloped males with large plumose antennae suited for pheromone detection (Tiemann, 1967; Crowson, 1972), and spectacular bicolor bioluminescence in Mastinocerini tribe larvae and females (Viviani and Bechara 1997). The railroad worms of the genus *Phrixothrix* are certainly the most spectacular examples of bioluminescence in Phengodidae, with red-light emitting cephalic and post cephalic lanterns and rows of yellow-green emitting lanterns along the body (Tiemann, 1967). The *Phrixothrix hirtus* has been used as an essential biological model for biochemical and molecular studies of luciferases and bioluminescence color modulation (Viviani et al., 1999; 2001; 2004; 2013; 2021; Viviani and Ohmiya, 2000; Amaral et al., 2016b; Bevilaqua et al., 2019). The bicolor bioluminescence is caused by the presence of paralogous luciferase isozymes in the head and lateral body lanterns (Viviani et al., 1999), which likely arose by events of gene duplication in the Mastinocerini tribe (Arnoldi et al., 2010). The luciferase cDNAs were cloned and characterized (Viviani et al., 1999), and structural–functional studies suggest that red bioluminescence is caused by a larger active site cavity (Bevilaqua et al., 2019). Recently, RNA-Seq analysis of the lateral lanterns and fat body of *P. hirtus* identified several luciferase-like enzymes in these tissues but only a luciferase in the photogenic tissues (Amaral et al., 2017a,b), suggesting that bioluminescence in phengodids did not arise from the fat body, such as in the case of fireflies (Viviani et al., 2008; Tonolli et al., 2011). With the exception of cDNA cloning, phylogeny studies based on molecular markers, and, more recently, functional genomics based on transcriptional analysis, the genome organization of this interesting family of beetles remains unknown.

Genome analysis has been generally used to understand genetic control and evolution, and in the case of the Elateroidea superfamily, to understand the processes that drive the diversification and evolution of bioluminescence. The genome size estimation in Elateroidea based on flow cytometry was also performed for all bioluminescent families (Hanrahan and Johnston, 2011; Liu et al., 2017; Lower et al., 2017). These studies indicated that the genome sizes among species of this superfamily range from 0.4 Gb to 2.2 Gb and identified a positive relationship between genome size and the number of transposable elements.

The first genome sequences and assemblies in Elateroidea were done for *Photinus pyralis* firefly (Lampyridae) and *Ignelater luminosus* click-beetle (Elateridae) (Fallon et al., 2018). These genomic analyses showed the divergence of the ancestral luciferase, supporting the independent origins of bioluminescence in Elateridae and Lampyridae. They also showed the existence of two distinct luciferase genes in *P. pyralis* firefly located in two chromosomes, suggesting events of gene duplication associated with a translocation between the chromosomes, explaining the presence of two luciferases isozymes in the lanterns and fat body in distinct life stages and tissues of fireflies (Strause and DeLuca, 1981; Viviani et al., 2008; Oba et al., 2010; Bessho-Uehara and Oba, 2017; Carvalho et al., 2020). In addition, genomic and transcriptomic analyses of two Palearctic firefly species, *Abscondita terminalis* and *Lamprigera yunnana*, suggested putative luciferin biosynthesis pathways in fireflies (Zhang et al., 2020), involving several gene products, which were also found in RNA-Seq analyses of other Elateroidea species (Vongsangnak et al., 2016; Amaral et al., 2017a,b; 2019b). Although such studies brought essential contributions and insights into the genome organization and evolution of the families Lampyridae and Elateridae, the genome information of the third family of Elateroidea, Phengodidae, remains unknown. The genome sequence and organization in this family is a missing link for better understanding the origin, evolution, and genetic control of biological processes such as paedomorphy and bioluminescence in this critical family and Elateroidea superfamily.

Here, we report the first draft genome assembly of a Phengodidae species, the South-American *P. hirtus* E.Olivier (1909) railroad worm, using both genomic short-read and mate-pair libraries. This species displays the largest genome among Elateriformia species studied, as well as the presence of several transposable element families. With this draft genome, we produced a partial genome assembly, which is a novel important source of information for future structural, genetic, biochemical, proteomic, and evolutionary studies and for comparative genomic analyses in Coleoptera species.

## 2. Materials and methods

### 2.1. Sampling, DNA extraction, and library construction

One larval individual of *P. hirtus* was manually collected at "Fazenda Santana" (Souza, Campinas/SP, Brazil, 22°44'45"S 47°06'33"W) and identified by Prof. Dr. Vadim Viviani. The sample was stored at −80 °C until DNA extraction. Genomic DNA was extracted from whole larvae using a DNeasy Blood Tissue Kit (Qiagen, USA) according to the manufacturer's instructions. The DNA quantity and quality were measured using a NanoVue (GE HealthCare) and Qubit 3.0 fluorometer (ThermoFisher, USA), and the integrity was checked in agarose gel. Two genomic DNA libraries were prepared: i.) using the TruSeq DNA PCR-free library prep kit with fragments of 150 bp and ii.) Nextera Mate Pair library prep with fragments of 2,000 bp (Illumina, USA). The short-read paired-end libraries were sequenced in two independent lanes, and the mate-pair library was sequenced in one lane using the Illumina HiSeq4000 platform (Illumina, USA). Library construction and sequencing were performed by Hokkaido System Science Co. (Sapporo, Hokkaido, Japan). The raw read data and final genome assembly of *P. hirtus* are deposited and available at the BioProject PRJNA741915.

### 2.2. Preprocessing data, de novo genome assembly, and annotation

The reads obtained were checked by FastQC v0.11.6 software (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and adaptors and low-quality reads (Phred Q $\leq$ 30) were removed using FASTX-TOOLKIT v0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/) for the paired-end library and NxTrim (O'Connell et al., 2015) for the mate-pair library. We used DeconSeq v.0.4.3 (Schmieder and Edwards, 2011) software and the RefSeq database (bacteria and viruses) (accessed in April 2018) to remove any contamination within the raw data sequencing. After filtering processes, we proceeded to downstream genomic assembly.

We estimated the best k-mer length for the *P. hirtus* genome assembly using KmerGenie 1.705 (Chikhi and Medvedev, 2014), which was k = 81 (data not shown). The genome coverage and size heterozygosity were estimated with the *estimate_genome_size.pl* script (available at https://github.com/josephryan/estimate_genome_size.pl), Jellyfish2 v.2.2.3 (Marçais and Kingsford, 2011) (parameters:*count -t 8 -C -m 21* –min-*quality = 20 –quality-start = 33*), CovEST (Hozza et al., 2015; available at https://github.com/mhozza/covest), and GenomeScope2 (Vurture et al., 2017; available at https://qb.cshl.edu/genomescope/genomescope2.0/). We used the *gatb-minia-pipeline* (available at https://github.com/GATB/gatb-minia-pipeline) for the *de novo* assembly using default settings and fixed the k-mer size to 81. We used the software RagTag v.2.1.0 (Alonge et al., 2021) to improve the genome assembly, employing the seven already sequenced genomes for Elateroidea as references (reference genome details in Table S1). The scaffolded statistics were evaluated using QUAST 5.0.0 (Gurevich et al., 2013). For the mitochondrial genome assembly, we used GetOrganelles software (Jin et al., 2020) with the default settings and the MITOS Web Server (Bernt et al, 2013) for annotation. Gene prediction was performed using the Maker pipeline v.3.01 (Cantarel et al., 2008), which included Augustus v.3.4.0 (Stanke et al., 2006), GeneMark-ES v.2.3c (Ter-Hovhannisyan et al., 2008), and FGENESH (Solovyev et al., 2006) gene prediction tools. We used the default settings combined with transcriptome-based nucleotide and amino acid sequences generated by Amaral et al. (2017a) to improve the prediction. Transposable element DNA elements were identified and annotated using RepeatModeler v.1.0.8 (see https://www.repeatmasker.org/RepeatModeler) and RepeatMasker

v.4.0.9 (see https://www.repeatmasker.org/). The annotation of the coding regions was conducted using the tool BLASTX against the SwissProt database (retrieved on 05/2020). The GO terms were plotted and visualized using WEGO2.0 (Ye et al., 2018). The completeness of the genes was estimated using BUSCO v.3.0.2 software using the Arthropoda database and the complete scaffold (Simão et al., 2015).

### 2.3. Orthologous protein clustering, gene family evolution, and phylogenetic analysis

Available genomic data from seven Elateriformia species (details in Table S1) were utilized to perform comparative genomic analyses. The orthogroup identification and the expanded and contracted gene families were identified with OrthoFinder v.2.0.9 (Emms and Kelly, 2015) and CAFÉ v.4.2.1, respectively, using birth and death rates (Han et al., 2013). The alignment was conducted in mafft v.4.87 software (Katoh et al., 2002). The ultrametric tree used to determine expansion and contraction family size was performed using r8s v.1.81 software (Sanderson, 2003). The phylogenetic relationship topology was estimated using 359 conserved single-copy orthologs in IQtree2 (Minh et al., 2020) in 1,000 ultrafast bootstraps. The divergence time estimates were performed using the phylogeny generated by IQtree2 and two secondary calibration points (the crown age of Elateroidea: 217.5–130.1 Mya, and the crown age of Lampyridae + Phengodidae; 123.3–90.8 Mya) obtained by Amaral et al. (2019a), within the treePL v.1.0 (Smith and O'Meara, 2012).

### 2.4. Positive selection genes (PSG) analysis

Positive selection genes analysis, on a genomic-scale, was carried out using the codeml program, a codon-based model within a maximum-likelihood framework, present in PAML v.4.8 package (Yang, 2007). This approach calculates the tree-based ratio between nonsynonymous (dN) and synonymous (dS) substitution, known as omega (ω). Thus, genes with the value of $\omega > 1$ evidenced positive selection. The codon-alignment was conducted in prank v.170427 software (Löytynoja, 2014). We also used a branch-site model to detect episodic positive selection at i.) *P. hirtus* branch and ii.) branches along the phylogenetic tree that lead to bioluminescent species. The Bayes Empirical Bayes (BEB) method was used to infer sites under positive selection and false positives were checked using the likelihood ratio test and multiple tests (FDR of 5 %) based on the R script proposed by Lee et al. (2017).

### 2.5. Luciferase gene identification

To identify the putative luciferase gene length, the raw reads were mapped against the luciferase of Phengodidae species available in public databases using the bowtie2 v.2.4.3 tool (Langmead and Salzberg, 2012). We processed the reads using Samtools v.1.9 (Li et al., 2009) and concatenated them using CAP3 v.10.2011 (Huang and Madan, 1999) software.

## 3. Results and discussion

### 3.1. De novo genome assembly and annotation

Here, we generated and assembled ~ 190 Gb (56.8-fold coverage) from Illumina short reads and ~ 65 Gb (20.2-fold coverage) from Illumina mate-pair reads (Table S2 and S3; scaffolded genome is detailed in Table S3). The genome of *P. hirtus* is the first genome in the Phengodidae family and the seventh sequenced and available genome within the Elateroidea superfamily. The assembly length, ~3.40 Gb, was consistent with the k-mer estimate genome size (~3.4 Gb in Jellyfish2.0 and CovEst, and ~ 3.35 Gb in GenomeScope; the k-mer distribution generated by Jellyfish2.0 is observed in Fig. S1). To date, this is the largest described genome among Elateroidea species (from 0.42 Gb to 2.2 Gb) and Coleoptera (from 0.15 Gb to 2.7 Gb) (Hanrahan and Johnston, 2011; Fu et al., 2017; Lower et al., 2017; Fallon et al., 2018). Previously, we showed that the mtDNA genome of *P. hirtus* is also larger than those of other Elateroidea species, with duplication events and a larger control region (Amaral et al., 2016a). Several hypotheses have been suggested and tested to evaluate the genome size correlation in Coleoptera, including body size (morphological; Palmer and Petitpierre, 1996), chromosome number (Petitpierre et al., 1993), methylation rate (Lechner et al., 2013), and reproductive fitness (Arnqvist et al., 2015). However, the genome size correlation in railroad worms needs to be better explored, including new species and morphological, environmental, and genomic data.

Gene prediction using the Maker pipeline shows that the best prediction was generated by Augustus, which resulted in 92,234 complete and partial gene products. The predicted proteins were searched against the SwissProt database and assigned to the putative function (Fig. S2). These observed GO terms are common in most arthropods and are responsible for basic physiologic processes and metabolic activities that are essential. The genome contains approximately 70 % complete single-copy orthologs and multicopy orthologs (BUSCO), which indicate that most parts of the genes were recovered. The percentage of observed GC content was ~ 32 %, and the estimated heterozygosity was approximately 0.25 %. We obtained a total of ~ 12 million transposable elements (2,326,217,477 bp, or 2.4 Gb), representing approximately 67 % of the total length of the assembled scaffolds (Table S4).

**Mitochondrial genome.** Recently, we published the partial sequence of *P. hirtus* mtDNA, with an incomplete portion of the *A + T-rich region* (Amaral et al., 2016a). Here, we could completely assemble the circular mitogenome of *P. hirtus*, which displayed 20,303 bp, similar to that previously studied in this species (Amaral et al., 2016a), confirming the largest mtDNA in Elateroidea. The mtDNA showed 2 ribosomal RNAs (rnaS and rnaL), 13 protein-coding genes (PCG), and 21 tRNAs (loss of tRNA-A; Table S5). The loss of tRNAs, such as *TrnD* in *Mesobuthus martensii* (Chelicerata), is commonly observed among Coleoptera (Gissi et al., 2008). A larger *A + T-rich region* (5,661 bp), which includes four partial copies of the *ND2* gene and tRNA-Q, was observed. Similar to the nuclear genome, these results also support the evidence of the high dynamism of the *P. hirtus* genome.

### 3.2. Repetitive element DNA content

The interspecies genome size variation could be the result of gene/genome duplication and/or deletion events (Blommaert, 2020). However, the rearrangements and duplication of transposable element (TEs) sequences may also imply larger genome sizes (Talla et al., 2017). Studies have shown the role of TE as potential substrates for new genes and their association with gene expression (e.g., epigenetic regulation), as well as a stress response (regulatory sequence) (Rech et al., 2019; Choi and Lee, 2020; Fedoroff, 2012). Thus, these elements could be major drivers of genome evolution in eukaryotes (Quesneville, 2020).

According to BUSCO, there was no excess of duplicated genes (7.42 % of duplicates). Nevertheless, the analysis of the landscape of TEs (Table S4) of *P. hirtus* showed that its genome has a great part of base-pair length composed of transposable elements (~67 %; 2.4 Gb). The percentages of TE content in Elateroidea vary from 19.8 % (~190 Mb) in *A. lateralis* to 42.6 % (~180 Mb) in *P. pyralis* (Fallon et al., 2018). In *P. hirtus*, the majority of these *TEs* were unclassified (46.83 %), followed by DNA transposons (12.3 %). Recent studies have also discussed the variation in TE contents among Coleoptera, Diptera (6 % in *Belgica antarctica* to 58 % in *Anopheles gambiae*), Hymenoptera (less than 6 % in *Apis mellifera* and *Athalia rosae*), and Orthoptera (58 % in *Locusta migratoria*), which suggested that transposable elements are highly dynamic in Insecta (Hazzouri et al., 2020).

### 3.3. Orthologous analysis and evolution of gene families in Elateriformia

The orthologous analysis using the amino acid sequences of Elateriformia species (*A. terminalis*, *L. yunnana*, *P. pyralis* (Lampyridae), *I. luminosus*, *Limonius californicus* (Elateridae), *P. hirtus* (Phengodidae), and *Agrilus planipennis* (Buprestidae) showed 17,862 orthologous gene families. Among them, 2,001 orthogroups (11.2 % of the total) were commonly shared among all species (Fig. 2), with only 359 orthogroups displaying a single gene copy, which were applied to phylogenetic reconstruction (Fig. 1a); 523 orthogroups were exclusively observed in the *P. hirtus* genome. The bioluminescent species used in this study (excluding *L. californicus* and *A. planipennis*) shared 2,477 orthogroups; however, only 440 displayed single-copy genes.

The molecular function ontology of the predicted gene shared among all Elateriformia species shows enrichment of catalytic activity and binding enzymes (Fig. S3). Among the gene products involved in enzymatic activities, the hydrolase and transferase classes were highlighted, while among the binding-related gene products, ion binding, heterocyclic binding, organic cyclic compound binding, and protein binding were highlighted (Table S6). The biological processes of the shared GO were mostly enriched for cellular processes (regulation of cellular process and cellular metabolic process), biological regulation (regulation of biological quality), regulation of biological process (regulation of metabolic process), and metabolic process (nitrogen compound metabolic process, organic substance metabolic process, and primary metabolic process) (Fig. S3A and S3C).

We separately evaluated the orthogroups associated only with the bioluminescent species of Elateroidea (Table S7). Considering these five species, we obtained 2,477 orthogroups, whereas 476 were exclusively for bioluminescent individuals. Here, we identified several families of general *odorant-binding proteins* (OBPs) responsible for recognizing and transporting hydrophobic odorants to the antennal sensilla and activating the olfactory signal transduction pathway (Li et al., 2016). In most flashing firefly species, such as *P. pyralis,* the bioluminescence signal pattern (flash, continuum/glow, etc.) is the main factor responsible for intraspecific long-distance communication. However, in railroad worms, it is known that pheromone detection plays a major role in sexual attraction (Jacobson, 2012). The presence of several OBP gene families supports the major role of pheromones in sexual attraction in Phengodidae. Thus, chemical and morphological studies using antennae and female-released pheromones are needed to better understand this communication system in phengodids.

The comparative analysis did not show clear differences in the presence or absence of gene families among bioluminescent species and their closely related non-bioluminescent species, with the exception of the OBP gene families in Phengodidae. These results suggest that both bioluminescent and non-bioluminescent species, in general, possess similar groups of gene families, including luciferase and/or luciferase-like genes, and genes associated with the luciferin biosynthetic pathway (Niwa et al., 2006; Oba et al., 2013; Amaral et al., 2017a; 2017b; 2019b; Zhang et al, 2020) such as sulfotransferases (Fallon et al., 2016), etc. Therefore, the transcriptional expression level of these genes may determine the spatial and temporal control of morphogenesis as well as the bioluminescence in Elateroidea, rather than genomic features. Future multi-omics approaches using transcriptomic, metabolomic, genomic, and proteomic data may help to support such hypotheses.

### 3.4. Analysis of unique P. hirtus orthogroups

We observed 523 orthogroups in *P. hirtus* that did not share direct orthology with the other studied species. In the annotation of these gene families, however, only a small part of the orthogroups was assigned (~12 %) (Table S8). From them, several TEs, such as LINE (long interspersed nuclear element), Mos1, ATP translocases, and PiggyBac, were found in these orthogroups. The high amount of these elements exclusively in the *P. hirtus* genome may partially explain the larger genome. The eukaryotic genome is highly dynamic and may have undergone events of gene duplication or even whole-genome duplications (Ting et al., 2004; Van de Peer et al., 2009; Mendivil and Ferrier, 2012). In arthropods, these mechanisms may have an important evolutionary significance when associated with adaptation to environmental changes and to the processes of biological and cellular regulation and simultaneous response to external stimuli (Kidwell, 2002; Chénais et al., 2012).

We also observed in *P. hirtus* the gene families of *craniofacial development protein* and *heat shock*, which seem to work in consonance with the development of wing dimorphism in arthropods (Carroll, 1995; Baral et al., 2019). However, in the case of *P. hirtus*, males undergo metamorphosis, and females do not. Moreover, for comparison between male and female morphology, RNA-Seq will be more informative than genome analysis, since we are not able to find sex-specific genes by genome comparison between males and females. Thus, more molecular and biochemical studies are necessary to understand the function of genes involved in metamorphosis and sexual differentiation in the family Phengodidae.

### 3.5. Expansion of orthologous gene families

The phylogenetic analysis showed a close relationship between Lampyridae and Phengodidae (Fig. 1a). Based on previous findings, Phengodidae diverged from the sister Palearctic family, Rhagophthalmidae, at approximately 73.4 Mya, and from Lampyridae at

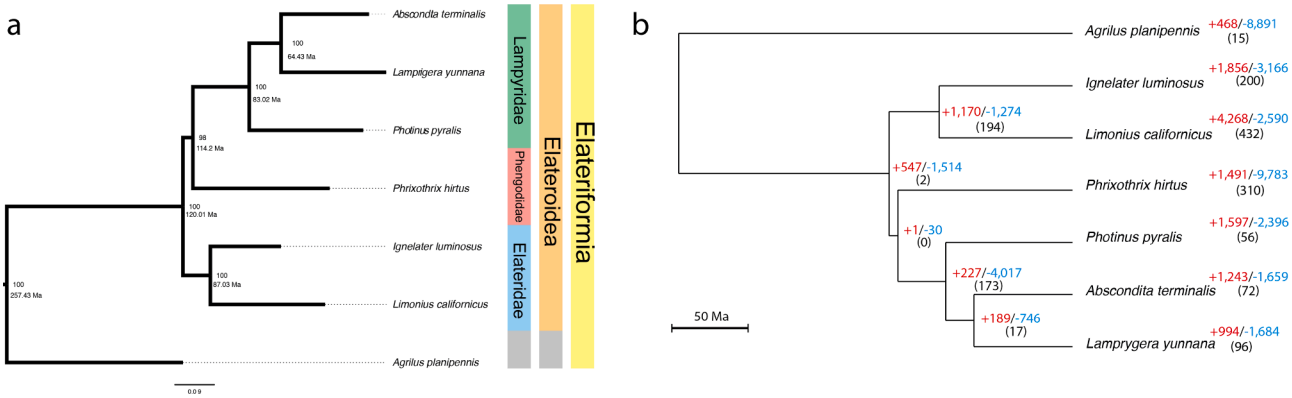


**Fig. 1.** Phylogenetic context of Elateriformia. a. Phylogenetic tree depicting the relationship within Elateroidea bioluminescent species, with *A. planipennis* (Buprestidae) as the outgroup. b. The summary tree shows the inference of the gene family evolution based on the expansion (+red) and contraction (- blue) gene families in Elateriformia. The number between parentheses represents the number of rapidly evolving gene families. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

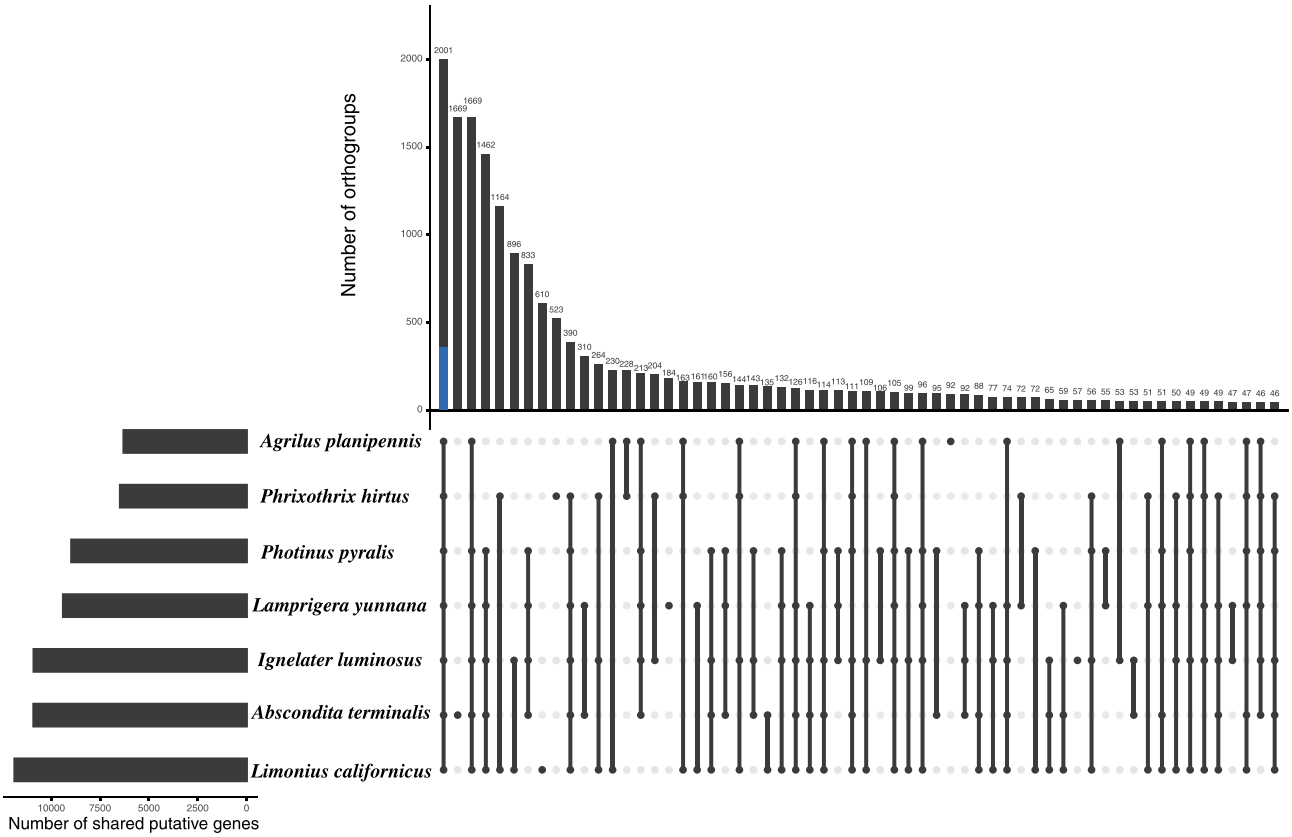


**Fig. 2.** Bar plot of shared and unique orthologs among Elateriformia genomes. In blue, the number of single-copy genes shared among the species in comparison to multiple copies orthogroups (in black). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

approximately 97.3 Ma (Amaral et al., 2019a). The ultrametric tree performed by the r8s software suggested that the divergence between Phengodidae and Lampyridae occurred even earlier, approximately 114 Mya. within the Early Cretacean.

In the past few years, comparative genomic studies have shown the dynamic aspect of genome size and gene families in rapidly evolving groups of species, such as plants and insects. The expansion and contraction of gene families seem to be pervasive and provide evidence that copy number changes are associated with natural selection acting under the particular adaptation of the species, such as changes in protein coding and regulatory regions (Han et al., 2009). These gene families with complex gene duplication histories in lineages deserve great attention. Here, we estimated the expansion and contraction of gene families among Elateriformia species.

The protein orthogroups obtained were managed to identify signatures of expansion in gene families among Elateriformia. The number of rapidly evolving, expanded, and contracted gene families is displayed in Fig. 1b. There were a total of 2,134 expanded and 7,581 contracted gene families among the internal branches of the Elateroidea families. The largest number of contraction gene families was found in *P. hirtus* (9,783), while the largest number of expansions was found in *Limonius* (4,268). The branches with the largest numbers of rapidly evolving gene families both led to the Elateridae and Lampyridae families (194 and 173, respectively). The ancestor branch of Elateroidea displayed 1,514 and 547 gene families with contractions and expansions, respectively.

In the main branch of Elateroidea, the expanded gene families are associated with detoxification/metabolism of xenobiotics (*cytochrome P450s*, *glutathione S-transferases, UDP-glucuronosyltransferases*, etc.) (Ahn et al., 2012; Zhu et al., 2016; Rane et al; 2019) and were widely identified among the transcriptome datasets of Elateroidea (Amaral et al., 2017a; 2017b; 2019b). These genes play fundamental roles in xenobiotic detoxification and the degradation of distinct molecules related to the insect diet, detoxification of metabolic compounds, resistance to pesticides, degradation of hormones (Xue et al., 2020), and the degradation of the lucibufagins (defensive steroids) present in some firefly species (McKinley and Lower, 2020). The expansion of the luciferase gene family is observed only in Elateridae, with the presence of several luciferases and luciferase-like enzymes, some of them which are also found in non-bioluminescent species. In Lampyridae, we observed the expansion of the *4-coumarate-CoA-like ligases*, but we did not identify the expansion of the luciferase gene family. However, we identified expansions of *superoxide dismutase* (SOD), *cystathionine-B-synthase,* and *cysteine-rich protein 2-binding*, which are associated with the protection of photocytes during hyperoxia and oxidative stress (Barros and Bechara, 1998), and with the availability of cysteine, the precursor of luciferin biosynthesis (Viviani et al., 2013; Kanie et al., 2016; Amaral et al., 2017a, 2017b; Zhang et al., 2020).

### 3.6. Gene families expansion and PSG in P. hirtus

From the 1,491 gene families expanded within *P. hirtus*, 1,346 were annotated (Table S9), most of them related to the multicellular organismal development, anatomical development and morphogenesis, and multiorganism processes (negative regulation of metabolic process). Besides, PSG analysis displayed 34 genes highly prone to adaptive selection involved with sex determination, organismal development, and morphogenesis (e.g., *doublesex- and mab-3-related transcription factor A2-like*, *tektin-2*, and *patj homolog*; Table 1; for more details on gene function enrichment see Table S10). Among them, are proteins that determine the formation, expression of spermatids, and apico-basal cell polarity (Amos, 2008; Sen et al., 2012; Zhuo et al., 2018), sex determination and differentiation, and cell-adhesion and morphogenetic role (Cabrera et al., 2017). Thus, these gene products are observed in insects associated with the ontogenetic evolution of the organism, orchestrating the

**Table 1**
List of positively selected genes and sites with posterior probabilities (model M1a and M2a) > 0.95(*) and > 0.99(**) to *P. hirtus* branch model analysis.

| Gene ID | lnL (alternative) | lnL (null) | ΔLRT | *p*-value | Significant BEB sites (Codon position, First species amino acid, Posterior probability) | Annotation |
|---|---|---|---|---|---|---|
| ENSG0010898 | −3532.24982 | −3536.806604 | 9.113568 | 0.002537205 | 224 F 0.992**;226 F 0.994**;226 F 0.994**;312 D 0.986*;313 S 0.998**;332 S 0.997** | *tribbles homolog 2* |
| ENSG0009695 | −9616.284029 | −9620.311544 | 8.05503 | 0.004537749 | 1027 E 0.982*;1089 V 0.993**;1117 R 0.979*;1118 A 0.994** | *tektin-2* |
| ENSG0009802 | −9612.657841 | −9615.710852 | 6.106022 | 0.013472202 | 1030 C 0.974*;1032 E 0.991**;1075 K 0.983* | *zinc finger MYND domain-containing protein 11* |
| ENSG0010630 | −9826.653828 | −9828.86049 | 4.413324 | 0.035659292 | 1059 A 0.965*;1061 N 0.965*;1062 V 0.987* | *patj homolog* |
| ENSG0009881 | −7514.638828 | −7516.817471 | 4.357286 | 0.03685097 | 1148 H 0.992** | *inhibitor of growth protein 3 isoform X2* |
| ENSG0009771 | −4768.982854 | −4772.220721 | 6.475734 | 0.010935714 | 124 A 0.961*;268 M 0.964*;518 S 0.992** | *ribosomal protein S6 kinase beta-1* |
| ENSG0009598 | −3433.038808 | −3435.887005 | 5.696394 | 0.016999804 | 136 H 1.000**;138 D 0.998**;139 S 0.993**;143 K 0.995**;173 K 0.994**;202 N 0.970*;287–0.998**;306–0.991** | *keratin. type I cytoskeletal 9-like* |
| ENSG0010825 | −3958.944895 | −3961.147891 | 4.405992 | 0.03581289 | 177 G 0.994**;223 V 0.991**;225 Q 0.998**;275 N 0.999**;290 S 0.954*;296 Q 0.982*;300 T 0.988* | *protein extra-macrochaeta* |
| ENSG0010568 | −17553.942 | −17557.75385 | 7.62371 | 0.005760586 | 1794 Q 0.996** | *inactive serine protease scarface/serine protease H164/proteoglycan 4-like* |
| ENSG0009519 | −4102.971156 | −4108.854656 | 11.767 | 0.000602901 | 191 S 1.000**;194 S 1.000**;232 H 1.000**;261 V 0.999**;265 N 0.990**;266 N 0.999**;271 K 0.991**;273 F 0.999**;278 H 0.999**;303 N 0.998**;305 Q 0.989*;307 V 0.970*;381 P 1.000**;398 E 0.999**;399 P 0.999** | *doublesex- and mab-3-related transcription factor A2-like* |
| ENSG0010360 | −7199.219248 | −7202.509468 | 6.58044 | 0.01031054 | 228 T 0.994**;376 S 0.990* | *actin-interacting protein 1* |
| ENSG0009515 | −4710.647248 | −4713.587461 | 5.880426 | 0.015310117 | 276 H 0.999** | *hypothetical protein FQR65_LT06461 [Abscondita terminalis]* |
| ENSG0011016 | −3866.798051 | −3868.803408 | 4.010714 | 0.045212 | 276 S 0.984* | *protein bric-a-brac 2-like* |
| ENSG0009844 | −5156.233372 | −5158.622022 | 4.7773 | 0.028837301 | 281 A 0.967* | *cuticle protein 18.7* |
| ENSG0010683 | −10912.90856 | −10916.80332 | 7.78951 | 0.005255045 | 296 A 0.999**;369 H 0.978*;385 M 0.969*;586 A 0.971*;591 K 0.999**;777 D 0.999**;778 P 0.998**;779 R 0.978* | *transcription elongation regulator 1* |
| ENSG0011001 | −5556.064378 | −5562.951982 | 13.775208 | 0.000206037 | 344 Q 0.995**;422 R 0.987*;424 Y 0.992**;436 A 0.992**;456 Q 0.974* | *cuticle protein 16.5. isoform A* |
| ENSG0009448 | −13752.44004 | −13754.41753 | 3.954984 | 0.046732762 | 373 Q 0.990*;667 V 0.981* | *flocculation protein FLO11-like* |
| ENSG0010619 | −2112.44771 | −2117.455663 | 10.015906 | 0.001551941 | 42 D 0.996**;73 F 0.999**;90 K 0.989*;95 K 0.995** | *hypothetical protein FQR65_LT14274 [Abscondita terminalis]* |
| ENSG0010394 | −6492.523227 | −6498.303029 | 11.559604 | 0.000674002 | 533 L 0.974*;534 S 0.996**;536 R 0.967*;537 C 1.000**;538 E 0.985* | *ruvB-like 2* |
| ENSG0010273 | −4098.001187 | −4100.29939 | 4.596406 | 0.032039054 | 59 R 0.963*;343 T 0.975* | *pre-mRNA-splicing regulator female-lethal(2) D* |
| ENSG0010779 | −3071.067155 | −3083.157105 | 24.1799 | 8.77433E-07 | 60 E 0.996**;65 F 0.971*;79 Y 0.997**;103 L 0.988*;106 A 0.987*;107 A 0.961*;120 S 0.950*;177 E 0.959*;178 D 0.982*;179–0.987*;213–0.991**;256 R 0.955* | *hypothetical protein FQR65_LT03434 [Abscondita terminalis]* |
| ENSG0010722 | −9890.522387 | −9893.80325 | 6.561726 | 0.010419526 | 685 Q 0.954*;752 Y 0.966*;756 E 0.993**;757 K 0.991**;758 G 0.995**;763 W 0.984*;819 G 0.995**;823 K 0.972*;846 E 0.979*;847 D 0.991** | *transcription factor TFIIIB component B'' homolog* |
| ENSG0010560 | −18191.01715 | −18193.56935 | 5.104412 | 0.023865067 | 796 N 0.988* | *nuclear receptor coactivator 1 isoform X3* |
| ENSG0009836 | −7523.011818 | −7528.740507 | 11.457378 | 0.000712107 | 82 K 0.977*;99 N 0.985*;302–0.981*;356 V 0.994**;670 K 0.997** | *DNA-binding protein D-ETS-6-like* |
| ENSG0010374 | −12916.96656 | −12919.01983 | 4.106526 | 0.042718025 | 908 E 0.982*;952 H 0.968* | *calpain-D-like* |

complex processes that contribute to embryogenic and growth development (Hill et al., 2010). Several gene families displayed distinct orthologs in *P. hirtus*, such as *axin-1 like* (anterior development in Coleoptera; Fu et al., 2012), *chromobox protein homolog* (crucial in the establishment and maintenance of heterochromatin in larvae; Shoji et al., 2013), *doublesex- and mab-3-related transcription factor* (Sex determination mechanisms and sex differentiation; Rather and Dhandare, 2019), and *forkhead box* (insulin signaling pathway and regulation of physiological processes and juvenile hormone degradation; Zeng et al., 2017). These gene products related to anatomical development and morphogenesis in *P. hirtus* might be associated with paedomorphic females and underdeveloped males. However, studies including specimens of both sexes and other omics approaches (e.g., RNA-Seq) could bring a better view of this question.

In the *P. hirtus* genome, we also identified olfactory, gustatory, odorant, and ionotropic receptor gene families in expansion, supporting the important role of intraspecific chemical communication in Phengodidae. The expansion of chemosensory gene families is associated with *gustatory* and *odorant receptor 22*, *olfactory receptor 2AG1*, *glutamate receptor subunit 1,* and *glutamate receptor ionotropic, kainate 2*. The chemosensory genes in insects are involved in mating, feeding, and coordinating actions (e.g., attack, defense, escape), among others (Yuvaraj et al., 2018; Blomquist and Ginzel, 2021), and are important physiological and ecological processes during the speciation process (mate isolation from its closely related species; Wu et al., 2019). Odorant and gustatory receptors are able to detect volatile chemicals, such as pheromones, which could be responsible for intraspecific communication, including mating. In Phengodidae, where the female is neotenic and the

winged male displays a well-developed antenna (Costa et al., 1999), chemosensory communication is critical for sexual attraction. However, we did not observe the expansion of the protein binding gene families, pheromones, or odorant carriers, indicating distinct evolutionary routes for the receptor and binding mechanisms in Phengodidae.

### 3.7. Evolution of bioluminescence in Phengodidae

Based on previous molecular analyses involving genomic and transcriptomic data and biochemical studies (Niwa et al., 2006; Viviani et al., 2013; Oba et al., 2013; Kanie et al., 2016; Vongsangnak et al., 2016; Amaral et al., 2017a; 2017b; 2019b; Fallon et al., 2018; Zhang et al., 2020), we searched for specifically described gene products that could be involved in the bioluminescence process in Elateroidea, mainly in *P. hirtus*. Here, we focused on luciferase evolution and putative candidates involved in bioluminescence emission control, antioxidant enzymes, and the luciferin biosynthesis pathway.

The number of AMP-forming enzymes, a gene family related to the activated acetate to acetyl-coenzyme A (e.g., fatty acyl-CoA synthetase) in which the luciferases are classified (Day et al., 2004), were abundant in the Elateroidea species (ca. 15), including bioluminescent and non-bioluminescent species. In the last few years, transcriptomic and genomic data were able to recover the distinct isoforms of luciferase and luciferase-like enzymes, and the comparison between their primary amino-acid sequences demonstrated the relationship between these isoforms and their evolution in Elateroidea. The presence of luciferase isozymes in the cephalic and lateral lanterns of Mastinocerini larvae was first shown by Viviani and Becham (1993) and confirmed after cDNA cloning of green and red-emitting luciferases (Viviani et al., 1999). Arnoldi et al. (2010), evaluating the relationship between the luciferase isoforms of Mastinocerini tribe species (Phengodidae), showed the presence of two luciferase paralogs, one in the cephalic lantern and one in the lateral body lanterns. These enzymes seem to be more closely related to the same lantern of different species than to the distinct lanterns of the same individual, suggesting an event of gene duplication and paralogy.

We were not able to completely assemble the genomic regions that contain the luciferase gene, including the putative promoter region (only partial fragments). Thus, the raw reads were mapped against known Phengodidae luciferases. Using this strategy, we identified two genes that displayed 90 % similarity to the luciferases of railroad worm species, which were named PhLuc1 and PhLuc2. The gene length for PhLuc1 was 2,019 bp (similar to that of yellow-green emitting lateral lantern luciferase), while that for PhLuc2 was 2,220 bp (similar to that of red-emitting cephalic lantern luciferase). Both genes comprise seven exons and six introns (Fig. 3), which is a similar number of intergenic components observed for the luciferase of *P. pyralis* firefly (Fallon et al., 2018) and *P. plagiophthalamus* click beetle (Elateridae; Feder and Velez, 2009). The possible duplication event that originated both putative luciferases (Arnoldi et al., 2010) could be followed by a dynamic structural genomic change, altering the intron size, mainly introns 1, 3, and 6 (Fig. 3). However, we did not observe evidence of intergenic combination in this species. The average sizes of exons and introns were ~ 230 bp and ~ 100 bp, respectively.

Recent studies using transcriptome and genomic data described possible gene products associated with bioluminescent control in Elateridae, Lampyridae, and Phengodidae (Amaral et al., 2017a; 2017b; 2019b; Zhang et al., 2020), such as *nitric oxide synthase* and *octopamine/dopamine receptors* (firefly flash control; Trimmer et al., 2001). Here, only *nitric oxide synthase* and *dopamine/octopamine receptor genes* were found. In Lampyridae genomes, only a copy of the nitric *oxide synthase* gene was also identified, suggesting a unique gene associated with the control of available oxygen concentration inside the photocytes, through inhibition of the respiratory chain. The number of *dopamine/octopamine receptors* in fireflies was between 8 and 10 copies, much higher than that observed in phengodids, consistent with the need for flash control in the adult stage of fireflies.

In the *P. hirtus* genome, gene products that were already described as potentially associated with luciferin biosynthesis in luminescent Elateroidea were identified. Among them, *adenosylhomocysteinase* and *cysteine sulfinic acid decarboxylase* are associated with the conversion of homocysteine to cysteine which may spontaneously react with *p*-benzoquinone to generate luciferin (Kanie et al., 2016). The gene products involved in tyrosine metabolism (*tyrosine aminotransferase, tyrosine hydroxylase, 4-hydroxyphenylpiruvate dioxygenase,* and *homogentisate 1,2-dioxygenase*) and the cascade reaction of the L-DOPA pathway (*dopamine/octopamine receptor, dopamine N-acetyltransferase, sodium-dependent dopamine transporter,* and *phenoloxidase* and *phenoloxidase activating factor*) were also observed. The *luciferin-regenerating enzyme* gene (LRE) and a gene of luciferin *sulfotransferase* (converting luciferin to a stable storage compound, *sulfoluciferin*; Fallon et al., 2016) were identified. However, we did not recover any *acyl-CoA thioesterase* (ACT) gene that could be involved in the conversion of *L-luciferin* into *D-luciferin* (Niwa et al., 2006), whereas we observed the expansion of this gene family among Lampyridae. Rather, we identified two genes of *palmitoyl-protein thioesterase* (a specific group within ACT), which is a lysosomal enzyme that removes fatty acyl groups from cysteine residues (Glaser et al., 2003), with a similar function to the ACT genes in peroxisomes (Lousa et al., 2013). These results suggest that luciferin biosynthesis may take similar pathways within Elateroidea bioluminescent species, although some steps or gene products involved in these pathways may have evolved differently during the diversification of the Elateroidea families, or were not recovered during genome annotation, such as the case of *Acyl-CoA thioesterase*.

## 4. Concluding remarks

We combined short and mate-pair reads generated with Illumina

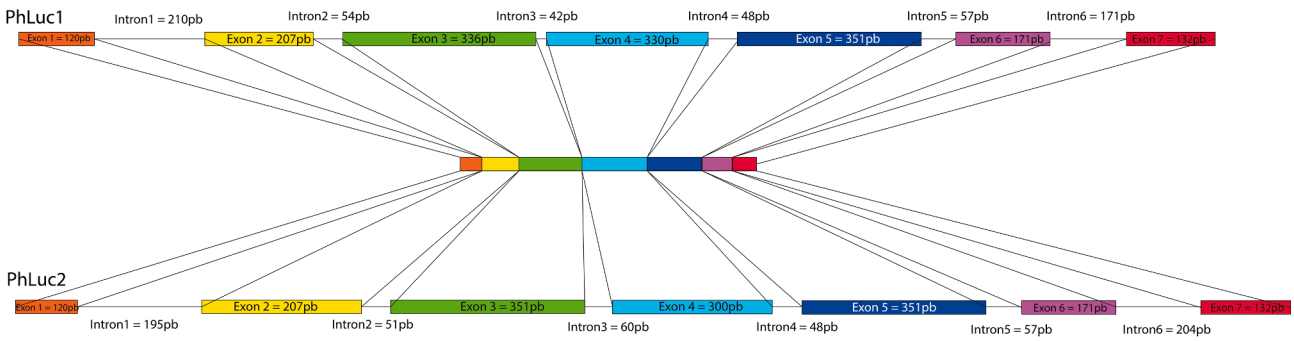


**Fig. 3.** Intergenic regions of the two putative luciferases of *P. hirtus*. The different box colors represent the seven exons identified in the sequences. PhLuc1 is more similar to the yellow-green emitting lateral lantern luciferase, while PhLuc2 is more similar to the red-emitting cephalic lantern luciferase. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

platform to assemble the first draft genome of a Phengodidae member, the South-American *P. hirtus* railroad worm. This is the largest genome observed in the Elateroidea superfamily, with >60 % of its size populated by TE, including the presence of several retrotransposable elements, such as LINE, MOS1, and PiggyBac. The comparative genomic analyses display PSGs and unique expanded gene families related to anatomical development, morphogenesis, and chemoreception, which may be consistent with the distinctive neoteny and sexual dimorphic development in this species, and with the intraspecific pheromone-mediated communication, typical of this family of beetles.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

*Acknowledgments*

We thank the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP grant 2010/05426-8 to VV; FAPESP n. 2017/207340 and 2014/20176-9 to DTA) for Financial Support.

*Author contributions*

VV, YO, YM, and DTA conceived the idea; DTA and IASB led both the analyses and manuscript writing; VV, YO, YM, and CR collaborated in manuscript writing. All authors contributed to the intellectual development of the paper, made multiple revisions, and approved the final draft.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2022.146917.

## References

Ahn, S.J., Vogel, H., Heckel, D.G., 2012. Comparative analysis of the UDP-glycosyltransferase multigene family in insects. Insect Biochem. Mol. Biol. 42 (2), 133–147.

Alonge, M., Kirsche, M., Aganezov, S., Wang, X., Lippman, X.B., Schatz, M.C., Soyk, S. 2021. Automated assembly scaffolding elevates a new tomato system for high-throughput genome editing. bioRxv.

Amaral, D.T., Mitani, Y., Ohmiya, Y., Viviani, V.R., 2016a. Organization and comparative analysis of the mitochondrial genomes of bioluminescent Elateroidea (Coleoptera: Polyphaga. Gene 586 (2), 254–262.

Amaral, D.T., Oliveira, G., Silva, J.R., Viviani, V.R., 2016b. A new orange emitting luciferase from the Southern-Amazon *Pyrophorus angustus* (Coleoptera: Elateridae) click-beetle: structure and bioluminescence color relationship, evolutional and ecological considerations. Photochem. Photobiol. Sci. 15 (9), 1148–1154.

Amaral, D.T., Silva, J.R., Viviani, V.R., 2017a. Transcriptional comparison of the photogenic and non-photogenic tissues of *Phrixothrix hirtus* (Coleoptera: Phengodidae) and non-luminescent *Chauliognathus flavipes* (Coleoptera: Cantharidae) give insights on the origin of lanterns in railroad worms. Gene Rep. 7, 78–86.

Amaral, D.T., Silva, J.R., Viviani, V.R., 2017b. Transcriptomes from the photogenic and non-photogenetic tissues and life stages of the *Aspisoma lineatum* firefly (Coleoptera: Lampyridae): Implications for the evolutionary origins of bioluminescence and its associated light organs. Gene Rep. 8, 150–159.

Amaral, D.T., Bonatelli, I.A., Cerri, R., Viviani, V.R., 2019a. Phylogenomic analyses and divergence time estimation of Elateroidea (Coleoptera) based on RNA-Seq data. Comp. Biochem. Physiol. - D: Genom. Proteom. 30, 283–289.

Amaral, D.T., Silva, J.R., Viviani, V.R., 2019b. RNA-Seq analysis of the bioluminescent and non-bioluminescent species of Elateridae (Coleoptera): Comparison to others photogenic and non-photogenic tissues of Elateroidea species. Comp. Biochem. Physiol. - D: Genom. Proteom. 29, 154–165.

Amos, L.A., 2008. The tektin family of microtubule-stabilizing proteins. Genome Biol. 9 (7), 1–8.

Arnoldi, F.G., da Silva Neto, A.J., Viviani, V.R., 2010. Molecular insights on the evolution of the lateral and head lantern luciferases and bioluminescence colors in Mastinocerini railroad-worms (Coleoptera: Phengodidae. Photochem. Photobiol. Sci. 9 (1), 87–92.

Arnqvist, G., Sayadi, A., Immonen, E., Hotzy, C., Rankin, D., Tuda, M., Johnston, J.S., 2015. Genome size correlates with reproductive fitness in seed beetles. Proc. Royal Soc. B 282 (1815), 20151421.

Baral, S., Arumugam, G., Deshmukh, R., Kunte, K., 2019. Genetic architecture and sex-specific selection govern modular, male-biased evolution of doublesex. Sci. Adv. 5 (5), eaau3753.

Barros, M.P., Bechara, E.J., 1998. Bioluminescence as a possible auxiliary oxygen detoxifying mechanism in elaterid larvae. Free Radi. Biol. Med. 24 (5), 767–777.

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Stadler, P.F., 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. 69 (2), 313–319.

Bessho-Uehara, M., Oba, Y., 2017. Identification and characterization of the Luc2-type luciferase in the Japanese firefly, *Luciola parvula*, involved in a dim luminescence in immobile stages. J. Lumin. 32 (6), 924–931.

Bevilaqua, V.R., Matsuhashi, T., Oliveira, G., Oliveira, P.S.L., Hirano, T., Viviani, V.R., 2019. *Phrixotrix* luciferase and 6′-aminoluciferins reveal a larger luciferin phenolate binding site and provide novel far-red combinations for bioimaging purposes. Sci. Rep. 9 (1), 1–17.

Blommaert, J., 2020. Genome size evolution: towards new model systems for old questions. Proc. Royal Soc. B 287 (1933), 20201441.

Blomquist, G.J., Ginzel, M.D., 2021. Chemical ecology, biochemistry, and molecular biology of insect hydrocarbons. Annu. Rev. Entomol. 66, 45–60.

Cabrera, G., Lundberg, U., Rodríguez-Ulloa, A., Herrera, M., Machado, W., Portela, M., Palomares, S., Espinosa, L.A., Ramos, Y., Durán, R., Besada, V., Vonasek, E., González, L.J., 2017. Protein content of the *Hylesia metabus* egg nest setae (Cramer [1775]) (Lepidoptera: Saturniidae) and its association with the parental investment for the reproductive success and lepidopterism. J. Proteom. 150, 183–200.

Cantarel, B.L., Korf, I., Robb, S.M., Parra, G., Ross, E., Moore, B., Yandell, M., 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 18 (1), 188–196.

Carroll, S.B., 1995. Homeotic genes and the evolution of arthropods and chordates. Nature 376 (6540), 479–485.

Carvalho, M.C., Tomazini, A., Amaral, D.T., Murakami, M.T., Viviani, V.R., 2020. Luciferase isozymes from the Brazilian *Aspisoma lineatum* (Lampyridae) firefly: origin of efficient pH-sensitive lantern luciferases from fat body pH-insensitive ancestors. Photochem. Photobiol. Sci. 19 (12), 1750–1764.

Chénais, B., Caruso, A., Hiard, S., Casse, N., 2012. The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. Gene 509 (1), 7–15.

Chikhi, R., Medvedev, P., 2014. Informed and automated k-mer size selection for genome assembly. Bioinformatics 30 (1), 31–37.

Choi, J.Y., Lee, Y.C.G., 2020. Double-edged sword: The evolutionary consequences of the epigenetic silencing of transposable elements. PLoS Genet. 16 (7), e1008872.

Costa, C., Vanin, S.A., Casari, S.A., Viviani, V.R., 1999. Larvae of Neotropical Coleoptera. XXVII. *Phrixotrix hirtus*: immatures, neotenic female, adult male and bionomic data (Phengodinae, Phengodidae. Coleoptera. Ilheringia. Serie Zoologia 86, 9–28.

Crowson, R.A., 1972. A review of the classification of Cantharoidea (Coleoptera), with the definition of two new families, Cneoglossidae and Omethidae. Revista de la Universidad de Madrid 21 (82), 35–77.

Day, J.C., Tisi, L.C., Bailey, M.J., 2004. Evolution of beetle bioluminescence: the origin of beetle luciferin. J. Lumin. 19 (1), 8–20.

Emms, D.M., Kelly, S., 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16 (1), 1–14.

Fallon, T.R., Li, F.S., Vicent, M.A., Weng, J.K., 2016. Sulfoluciferin is biosynthesized by a specialized luciferin sulfotransferase in fireflies. Biochemistry 55 (24), 3341–3344.

Fallon, T.R., Lower, S.E., Chang, C.H., Bessho-Uehara, M., Martin, G.J., Bewick, A.J., Weng, J.K., 2018. Firefly genomes illuminate parallel origins of bioluminescence in beetles. Elife 7, e36495.

Feder, J.L., Velez, S., 2009. Intergenic exchange, geographic isolation, and the evolution of bioluminescent color for *Pyrophorus* click beetles. Evol. 63 (5), 1203–1216.

Fedoroff, N.V., 2012. Transposable elements, epigenetics, and genome evolution. Science 338 (6108), 758–767.

Fu, X., Li, J., Tian, Y., Quan, W., Zhang, S., Liu, Q., Hu, J., 2017. Long-read sequence assembly of the firefly *Pyrocoelia pectoralis* genome. GigaScience 6 (12), gix112.

Fu, J., Posnien, N., Bolognesi, R., Fischer, T.D., Rayl, P., Oberhofer, G., Bucher, G., 2012. Asymmetrically expressed axin required for anterior development in *Tribolium*. PNAS 109 (20), 7782–7786.

Gissi, C., Iannelli, F., Pesole, G., 2008. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity 101 (4), 301–320.

Glaser, R.L., Hickey, A.J., Chotkowski, H.L., Chu-LaGraff, Q., 2003. Characterization of Drosophila palmitoyl-protein thioesterase 1. Gene 312, 271–279.

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29 (8), 1072–1075.

Han, M.V., Demuth, J.P., McGrath, C.L., Casola, C., Hahn, M.W., 2009. Adaptive evolution of young gene duplicates in mammals. Genome Res. 19 (5), 859–867.

Han, M.V., Thomas, G.W., Lugo-Martinez, J., Hahn, M.W., 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Mol. Biol. Evol. 30 (8), 1987–1997.

Hanrahan, S.J., Johnston, J.S., 2011. New genome size estimates of 134 species of arthropods. Chromosome Res. 19 (6), 809–823.

Hazzouri, K.M., Sudalaimuthuasari, N., Kundu, B., Nelson, D., Al-Deeb, M.A., Le Mansour, A., Amiri, K.M., 2020. The genome of pest *Rhynchophorus ferrugineus* reveals gene families important at the plant-beetle interface. Commun. Biol. 3 (1), 1–14.
Hill, D.P., Berardini, T.Z., Howe, D.G., Van Auken, K.M., 2010. Representing ontogeny through ontology: a developmental biologist's guide to the gene ontology. Mol. Reproduct. Dev.: Incorporating Gamete Res. 77 (4), 314–329.
Hozza, M., Vinař, T., Brejová, B., 2015. In: How Big Is That Genome? Estimating Genome Size and Coverage From K-mer Abundance Spectra. Springer International Publishing, New York, pp. 199–209.
Huang, X., Madan, A., 1999. CAP3: A DNA sequence assembly program. Genome Res. 9 (9), 868–877.
Jacobson, M., 2012. Insect sex pheromones. Elsevier.
Jin, J.J., Yu, W.B., Yang, J.B., Song, Y., Depamphilis, C.W., Yi, T.S., Li, D.Z., 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 21 (1), 1–31.
Kanie, S., Nishikawa, T., Ojika, M., Oba, Y., 2016. One-pot non-enzymatic formation of firefly luciferin in a neutral buffer from p-benzoquinone and cysteine. Sci. Rep. 6, 24794.
Katoh, K., Misawa, K., Kuma, K.I., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30 (14), 3059–3066.
Kidwell, M.G., 2002. Transposable elements and the evolution of genome size in eukaryotes. Genetica 115 (1), 49–63.
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9 (4), 357.
Lechner, M., Marz, M., Ihling, C., Sinz, A., Stadler, P.F., Krauss, V., 2013. The correlation of genome size and DNA methylation rate in metazoans. Theory in Biosci. 132 (1), 47–60.
Lee, R., Wiel, L., van Dam, T.J.P., Huynen, M.A., 2017. Genome-scale detection of positive selection in nine primates predicts human-virus evolutionary conflicts. Nucleic Acids Res. 45, 10634–10638.
Li, G., Chen, X., Li, B., Zhang, G., Li, Y., Wu, J. 2016. Binding properties of general odorant binding proteins from the oriental fruit moth, *Grapholita molesta* (Busck) (Lepidoptera: Tortricidae. PLoS One 11(5), e0155096.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25 (16), 2078–2079.
Liu, G.C., Dong, Z.W., He, J.W., Zhao, R.P., Wang, W., Li, X.Y., 2017. Genome size of 14 species of fireflies (Insecta, Coleoptera Lampyridae. Zoological Res. 38 (6), 449.
Lousa, C.D.M., van Roermund, C.W., Postis, V.L., Dietrich, D., Kerr, I.D., Wanders, R.J., Theodoulou, F.L., 2013. Intrinsic acyl-CoA thioesterase activity of a peroxisomal ATP binding cassette transporter is required for transport and metabolism of fatty acids. PNAS 110 (4), 1279–1284.
Lower, S.S., Johnston, J.S., Stanger-Hall, K.F., Hjelmen, C.E., Hanrahan, S.J., Korunes, K., Hall, D., 2017. Genome size in North American fireflies: substantial variation likely driven by neutral processes. Genome Biol. Evol. 9 (6), 1499–1512.
Löytynoja, A., 2014. Phylogeny-aware alignment with PRANK. Methods Mol. Biol. 1079, 155–170.
Marçais, G., Kingsford, C., 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27 (6), 764–770.
McKinley, C.N., Lower, S.E., 2020. Comparative transcriptomics reveals gene families associated with predatory behavior in *Photuris* femme fatale fireflies. Genes 11 (6), 627.
Mendivil, O.R., Ferrier, D.E.K., 2012. Mechanisms of gene duplication and translocation and progress towards understanding their relative contributions to animal genome evolution. Int. J. Evol. Biol. 2012, 1–10.
Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37 (5), 1530–1534.
Niwa, K., Nakamura, M., Ohmiya, Y., 2006. Stereoisomeric bio-inversion key to biosynthesis of firefly D-luciferin. FEBS Lett. 580 (22), 5283–5287.
O'Connell, J., Schulz-Trieglaff, O., Carlson, E., Hims, M.M., Gormley, N.A., Cox, A.J., 2015. NxTrim: optimized trimming of Illumina mate pair reads. Bioinformatics 31 (12), 2035–2037.
O'Keefe, S.T., Arnett, R.H., Thomas, M.C., Skelley, P.E., Frank, J.H., 2002. Phengodidae LeConte 1861. CRC Press LLC, Boca Raton, pp. 181–186.
Oba, Y., Mori, N., Yoshida, M., Inouye, S., 2010. Identification and characterization of a luciferase isotype in the Japanese firefly, *Luciola cruciata*, involving in the dim glow of firefly eggs. Biochemistry 49 (51), 10788–10795.
Oba, Y., Yoshida, N., Kanie, S., Ojika, M., Inouye, S. 2013. 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 8(12), e84023.
Palmer, M., Petitpierre, E., 1996. Relationship of genome size to body size in *Phylan semicostatus* (Coleoptera: Tenebrionidae. Ann. Entomol. Soc. Am. 89 (2), 221–225.
Petitpierre, E., Segarra, C., Juan, C., 1993. Genome size and chromosomal evolution in leaf beetles (Coleoptera. Chrysomelidae. Hereditas 119 (1), 1–6.
Quesneville, H., 2020. Twenty years of transposable element analysis in the *Arabidopsis thaliana* genome. Mob. DNA 11 (1), 1–13.
Rane, R.V., Ghodke, A.B., Hoffmann, A.A., Edwards, O.R., Walsh, T.K., Oakeshott, J.G., 2019. Detoxifying enzyme complements and host use phenotypes in 160 insect species. Curr. Opin. Insect Sci. 31, 131–138.
Rather, M.A., Dhandare, B.C., 2019. Genome-Wide identification of doublesex and Mab-3-Related transcription factor (DMRT) genes in Nile tilapia (*Oreochromis niloticus*. Biotechnol. Rep, 24, e00398.
Rech, G.E., Bogaerts-Marquez, M., Barrón, M.G., Merenciano, M., Villanueva-Cañas, J.L., Horváth, V., González, J., 2019. Stress response, behavior, and development are shaped by transposable element-induced mutations in Drosophila. PLoS Genet. 15 (2), e1007900.
Sanderson, M.J., 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19 (2), 301–302.
Schmieder, R., Edwards, R., 2011. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. PLoS ONE 6 (3), e17288.
Sen, A., Nagy-Zsvér-Vadas, Z., Krahn, M.P., 2012. Drosophila PATJ supports adherens junction stability by modulating Myosin light chain activity. J. Cell Biol. 199 (4), 685–698.
Shoji, K., Kiuchi, T., Hara, K., Kawamoto, M., Kawaoka, S., Arimura, S.I., Katsuma, S., 2013. Characterization of a novel chromodomain-containing gene from the silkworm, Bombyx mori. Gene 527 (2), 649–654.
Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31 (19), 3210–3212.
Smith, S.A., O'Meara, B.C., 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. Bioinformatics 28 (20), 2689–2690.
Solovyev, V., Kosarev, P., Seledsov, I., Vorobyev, D., 2006. Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biol. 7 (1), 1–12.
Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., Morgenstern, B., 2006. AUGUSTUS: *ab initio* prediction of alternative transcripts. Nucleic Acids Res. 34 (suppl_2), W435–W439.
Strause, L.G., DeLuca, M., 1981. Characteristics of luciferases from a variety of firefly species: evidence for the presence of luciferase isozymes. Insect Biochem. 11 (4), 417–422.
Talla, V., Suh, A., Kalsoom, F., Dincă, V., Vila, R., Friberg, M., Backström, N., 2017. Rapid increase in genome size as a consequence of transposable element hyperactivity in wood-white (*Leptidea*) butterflies. Genome Biol. Evol. 9 (10), 2491–2505.
Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y.O., Borodovsky, M., 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res. 18 (12), 1979–1990.
Ting, C. T., Tsaur, S. C., Sun, S., Browne, W. E., Chen, Y. C., Patel, N. H. Wu, C.I. 2004. Gene duplication and speciation in Drosophila: Evidence from the *Odysseus locus*. Proc. Natl. Acad. Sci. USA 101, 12232–12235.
Tiemann, D.L., 1967. Observation on the natural history of the western banded glowworm Zarhipis integripennis (LeConte) (Coleoptera: Phengodidae). In: Proceedings of the California Academy of Sciences, 4th Series, vol. 35(12), pp. 235–264.
Tonolli, P.N., Okawachi, F.M., Abdalla, F.C., Viviani, V.R., 2011. Bioluminescent fat body of larval *Aspisoma lineatum* (Coleoptera: Lampyridae) firefly: ontogenic precursor of lantern's photogenic tissue. Ann. Entomol. Soc. Am. 104 (4), 761–767.
Trimmer, B.A., Aprille, J.R., Dudzinski, D.M., Lagace, C.J., Lewis, S.M., Michel, T., Zayas, R.M., 2001. Nitric oxide and the control of firefly flashing. Science 292 (5526), 2486–2488.
Van de Peer, Y., Maere, S., Meyer, A., 2009. The evolutionary significance of ancient genome duplications. Nat. Rev. Genet. 10, 725–732.
Viviani, V.R., 2002. The origin, diversity, and structure function relationships of insect luciferases. Cell. Mol. Life Sci. 59 (11), 1833–1850.
Viviani, V.R., Becham, E.J., 1993. Biophysical and biochemical aspects of phengodid (railroad-worm) bioluminescence. Photochem. Photobiol. 58 (4), 615–622.
Viviani, V.R., Bechara, E.J., 1997. Bioluminescence and biological aspects of Brazilian railroad-worms (Coleoptera: Phengodidae. Ann. Entomol. Soc. Am. 90 (3), 389–398.
Viviani, V.R., da Silva, J., Neto, A., Ohmiya, Y., 2004. The influence of the region between residues 220 and 344 and beyond in *Phrixotrix* railroad worm luciferases green and red bioluminescence. Protein Eng. Desi. Sel. 17 (2), 113–117.
Viviani, V.R., Okawachi, F.M., Scorsato, V., Abdalla, F.C., 2008. CCD imaging of basal bioluminescence in larval fireflies: clues on the anatomic origin and evolution of bioluminescence. Photochem. Photobiol. Sci. 7 (4), 448–452.
Viviani, V.R., Carmargo, I.A., Amaral, D.T., 2013. A transcriptional survey of the cDNA library of *Macrolampis* sp2 firefly lanterns (Coleoptera: Lampyridae. Comp. Biochem. Physiol. - D: Genom. Proteom. 8 (1), 82–85.
Viviani, V.R., Bechara, E.J., Ohmiya, Y., 1999. Cloning, sequence analysis, and expression of active *Phrixothrix* railroad-worms luciferases: relationship between bioluminescence spectra and primary structures. Biochemistry 38 (26), 8271–8279.
Viviani, V.R., Bevilaqua, V.R., de Souza, D.R., Pelentir, G.F., Kakiuchi, M., Hirano, T., 2021. A very bright far-red bioluminescence emitting combination based on engineered railroad worm luciferase and 6′-Amino-Analogs for bioimaging purposes. Int. J. Mol. Sci. 22 (1), 303.
Viviani, V.R., Ohmiya, Y., 2000. Bioluminescence color determinants of *Phrixothrix* railroad-worm luciferases: chimeric luciferases, site-directed mutagenesis of Arg 215 and guanidine effect. Photochem. Photobiol. 72 (2), 267–271.
Viviani, V., Uchida, A., Suenaga, N., Ryufuku, M., Ohmiya, Y., 2001. Thr226 is a key residue for bioluminescence spectra determination in beetle luciferases. Biochem. Biophys. Res. Commun. 280 (5), 1286–1291.
Vongsangnak, W., Chumnanpuen, P., Sriboonlert, A., 2016. Transcriptome analysis reveals candidate genes involved in luciferin metabolism in *Luciola aquatilis* (Coleoptera: Lampyridae. PeerJ 4, e2534.
Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., Schatz, M.C., 2017. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics 33 (14), 2202–2204.
Wittmer, W. 1976. Arbeiten zu einer Revision der Familie Phengodidae (Coloptera. Entomol Arb Mus G Frey Tutzing Munch.

Wu, Y., Liu, Y.Y., Chen, X.S., 2019. Whole-genome identification of chemosensory genes in two sister blister beetles provides insights into chemosensory evolution. Research Square.
Xue, H.J., Niu, Y.W., Segraves, K.A., Nie, R.E., Hao, Y.J., Zhang, L.L., Yang, X.K., 2020. The draft genome of the specialist flea beetle *Altica viridicyanea* (Coleoptera: Chrysomelidae) provides insight into host plant specialization. BMC Genom. 22 (1), 1–18.
Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24 (8), 1586–1591.
Ye, J., Zhang, Y., Cui, H., Liu, J., Wu, Y., Cheng, Y., Shi, C., 2018. WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update. Nucleic Acids Res. 46 (W1), W71–W75.
Yuvaraj, J.K., Andersson, M.N., Zhang, D.D., Löfstedt, C., 2018. Antennal transcriptome analysis of the chemosensory gene families from Trichoptera and basal Lepidoptera. Front. Physiol. 9, 1365.
Zeng, B., Huang, Y., Xu, J., Shiotsuki, T., Bai, H., Palli, S.R., Tan, A., 2017. The FOXO transcription factor controls insect growth and development by regulating juvenile hormone degradation in the silkworm, *Bombyx mori*. J. Biol. Chem. 292 (28), 11659–11669.
Zhang, R., He, J., Dong, Z., Liu, G., Yin, Y., Zhang, X., Li, X., 2020. Genomic and experimental data provide new insights into luciferin biosynthesis and bioluminescence evolution in fireflies. Sci. Rep. 10 (1), 1–19.
Zhu, F., Moural, T.W., Nelson, D.R., Palli, S.R., 2016. A specialist herbivore pest adaptation to xenobiotics through up-regulation of multiple Cytochrome P450s. Sci. Rep. 6 (1), 1–10.
Zhuo, J.C., Hu, Q.L., Zhang, H.H., Zhang, M.Q., Jo, S.B., Zhang, C.X., 2018. Identification and functional analysis of the doublesex gene in the sexual development of a hemimetabolous insect, the brown planthopper. Insect Biochem. Mol. Biol. 102, 31–42.

## Further reading

Viviani, V.R., Bechara, E.J., 1995. Bioluminescence of Brazilian fireflies (Coleoptera: Lampyridae): spectral distribution and pH effect on luciferase-elicited colors. Comparison with elaterid and phengodid luciferases. Photochem. Photobiol. 62 (3), 490–495.