scientific reports

Check for updates

OPEN

Absence of the luciferase gene in the genome of the kleptoprotein bioluminescent fish *Parapriacanthus ransonneti*

Manabu Bessho-Uehara[1,2]✉, Katsushi Yamaguchi[3], Keita Koeda[4], Shohei Matsuzaki[5], Taro Maeda[6] & Shuji Shigenobu[3]

The metabolic and physiological functions of the living organisms are primarily governed by enzymes encoded within their own genomes. An exception to this doctrine is the phenomenon of kleptobiology, where individuals sequester and use functional modules from their prey. The fish *Parapriacanthus ransonneti* possesses a luciferase that catalyzes a light-emitting reaction in its light organs. Intriguingly, the protein sequence of this luciferase is identical to that of the ostracod-derived luciferase. We recently demonstrated that the fish sequesters and retains the luciferase from the luminous prey, leading to the hypothesis that the fish does not have a gene encoding the luciferase. To test this hypothesis, here, we produced a high-quality draft genome of *P. ransonneti*. The assembled size of 625 Mbp closely matches the experimentally estimated genome size of 613 Mbp, and we constructed a gene model with a high BUSCO score. Our search within this gene model did not detect any ostracod-type luciferase genes. Expanding our search to include the entire assembled genome, raw reads, and the transcriptome also supported the absence of the luciferase genes. These results suggest that *P. ransonneti* lacks the luciferase gene, supporting our previous discovery that it utilizes stolen proteins for bioluminescence.

Keywords Bioluminescence, Kleptoprotein, Kleptobiology, *Cypridina noctiluca*, Horizontal gene transfer

Biology has long focused on understanding how genes determine the traits of the organism, following the central dogma of molecular biology[1]. In most cases, a phenotype directly depends on its genotype[2–4]. However, there are exceptions where traits are influenced by interactions with other organisms[5,6]. For example, in symbiotic relationships like endosymbiosis, two organisms can complement each other's genetic and phenotypic gaps[7].

Notably, some organisms can use components of their prey to extend their traits. This process of sequestration allows them to incorporate and use functional components from their food. For instance, aeolid nudibranchs sequester stinging cell components (nematocysts) from their prey and store them in the appendages[8–10]. The sacoglossan sea slugs retain chloroplasts from algae and take advantage of photosynthesis[11]. These sequestered organelles, called kleptocnidae and kleptoplasts, respectively, retain their original functions for months. These components remain functional and benefit the host, even though the host lacks the genes to produce them[10,12].

A similar process was recently found in the luminous fish *Parapriacanthus ransonneti* (Pempheridae, Acropomatiformes, Teleostei), which acquires its bioluminescence by taking a luminescent protein from bioluminescent ostracods *Cypridina noctiluca* (Fig. 1)[13]. *Parapriacanthus ransonneti* possesses a luciferase—an enzyme that catalyzes a light-emitting reaction—with an amino acid sequence identical to that of the ostracod *C. noctiluca*[13]. Despite using genomic PCR, reverse transcription PCR (RT-PCR), and cell-free expression systems, we were unable to detect the gene encoding this protein in the fish[13]. The amount of the luciferase activity in *P. ransonneti* decreased when the fish specimens were cultured for several months without feeding the ostracods[13]. The hypothesis of sequestration of ingested protein, kleptoprotein, was further demonstrated by feeding

[1]Frontier Research Institute for Interdisciplinary Sciences, Tohoku University, 6-3 Aramaki, Aoba, Sendai, Miyagi 980-8578, Japan. [2]Graduate School of Life Sciences, Tohoku University, 6-3 Aramaki, Aoba, Sendai, Miyagi 980-8578, Japan. [3]Trans-Omics Facility, National Institute for Basic Biology, Nishigonaka 38, Myodaiji, Okazaki 444-8585, Japan. [4]Faculty of Science, University of the Ryukyus, 1 Senbaru, Nishihara, Okinawa, Japan. [5]Okinawa Churashima Foundation, Kunigami-gun, Aza Ishikawa 888, Motobu-cho, Okinawa 905-0206, Japan. [6]Institute for Advanced Biosciences, Keio University, Kakuganji 246-2, Mizukami, Tsuruoka, Yamagata 997-0052, Japan. ✉email: bessho.manabu.lumi@gmail.com

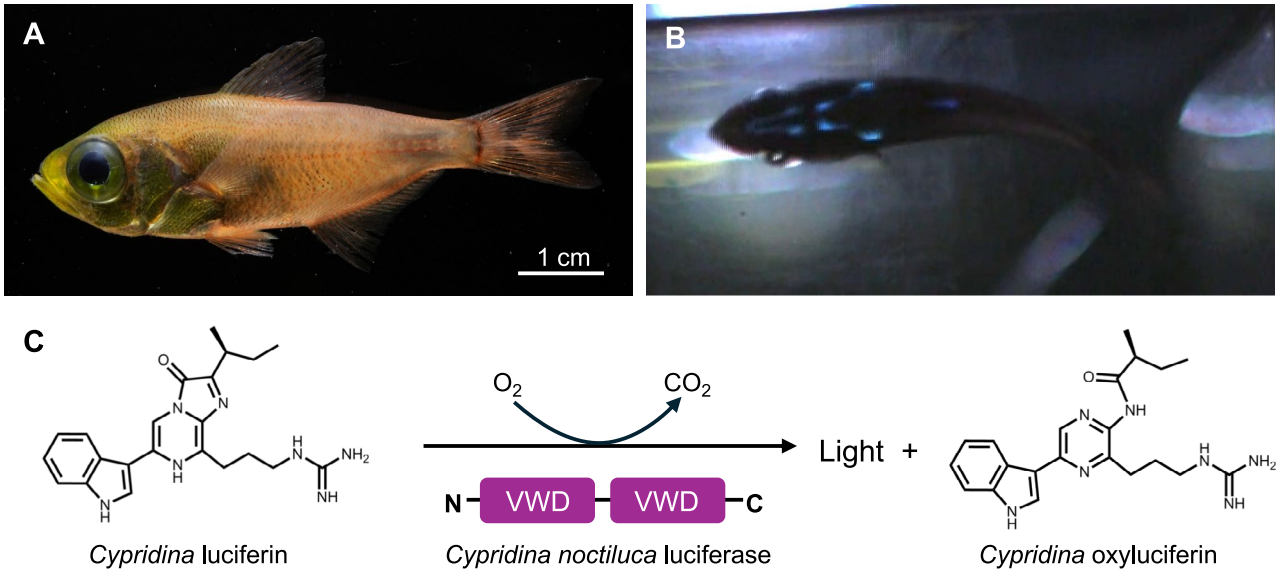

Fig. 1. Bioluminescence of *Parapriacanthus ransonneti*. (**A**) Lateral view of *Parapriacanthus ransonneti*. (**B**) Bioluminescence observed from the ventral view in an aquarium tank under dim light conditions. (**C**) Schematic representation of the bioluminescence reaction. The light producing reaction is characterized by the oxidation of *Cypridina* luciferin catalyzed by *Cypridina noctiluca* luciferase. The luciferase is characterized by a domain architecture consisting of two von Willebrand factor type D (VWD) domains.

experiments[13]. Based on these results, we hypothesized that the luciferase gene is not present in the genome of *P. ransonneti* and that the protein is sequestered from ingested prey, a process termed kleptoproteinism[13]. However, no genome-wide analysis of *P. ransonneti* had been performed, so it remained uncertain whether the luciferase gene was truly absent in its genome. Clarifying this is key to understanding how organisms can use proteins from their prey instead of producing them on their own.

In this study, we analyzed the genome of *P. ransonneti* using high-precision sequencing technology. We searched for the ostracod luciferase genes in the assembled genome, raw sequencing data, and annotated gene models, but found no trace of them. These results confirm that *P. ransonneti* does not have the luciferase gene and relies entirely on sequestration for its bioluminescence.

Materials and methods

Ethics approval

The animal experiments in this study were approved by the Institutional Animal Experiment Committee of Nagoya University and were conducted in accordance with the Regulations on Animal Experiments at the university (Approval Nos. S210622, S220019, and S230010). All experimental procedures were designed to minimize animal suffering and were performed in compliance with relevant institutional and national guidelines. Furthermore, the authors confirm that all experiments were performed and reported in compliance with the ARRIVE guidelines.

Fish specimen

The live specimens of *P. ransonneti* were provided by Okinawa Churaumi Aquarium. The fish were originally collected by local fishermen from near Okinawa Island in July 2021 and subsequently supplied to the aquarium. The specimens were fed minced non-luminous fishes or mysids but not luminous ostracods, such as *Cypridina noctiluca* nor *Vargula hilgendorfii,* for 15 months. This feeding condition minimizes the risk of contamination of the genomic DNA of the luminous ostracod which possess cypridinid-type luciferases. The specimens were anesthetized with tricaine methanesulfonate (Merck, Darmstadt, Germany), also known as MS-222, combined with ice prior to dissection for tissue collection.

For de novo mRNA sequencing, muscle tissue was collected from *Pempheris nyctereutes* (Pempheridae) and *Glaucosoma buergeri* (Glaucosomatidae), obtained from a local fish market in Daxi Fishing Port and the Keelung Night Fish Market, respectively, collected from off the northeastern coast of Taiwan in March 2023.

Genome size determination

We determined the genome size of *P. ransonneti* by fluorescence-activated cell sorting (FACS) analysis. The frozen whole brain was minced by using a blade in 1 mL of phosphate-buffered saline (PBS). The minced solution was mixed with 1 µL of 10% triton-X (Merck) and 4 µL of 100 mg/mL RNase A (QIAGEN, Germany), followed by filtration using a cell strainer CellTrics (Sysmex, USA) with 30 µm mesh. The nuclei were stained by adding 20 µL of 0.5 mg/mL Propidium Iodide Solution (PI) (SONY, Japan). The nuclei size was analyzed by using SH800

Cell Sorter (SONY). Three technical replicates of this sample were performed. Independent runs for extracted fruit fly nuclei (*Drosophila melanogaster,* 173.3 Mbp) were performed as calibration standards.

Genome sequencing

We constructed the PacBio Sequel HiFi long-read library from *P. ransonneti.* To reduce the contamination risk of environmental DNA, the dorsal muscle tissue was isolated without including skin. The Genomic DNA was extracted from 316 mg of frozen tissue using Blood and Cell Culture DNA kits (QIAGEN) containing beta-mercaptoethanol, Proteinase K, and RNase. The extracted DNA was purified by Genomic-tip 20/G (QIAGEN) according to the manufacturer's protocol and eluted with 2 mL of solution. The DNA was further purified with AMPureXP and eluted in 200 μL. The quality of the purified DNA was checked with the dsDNA HS Assay kit and the RNA HS Assay kits. The RNA contamination was lower than the detection limit (lower than 4 ng/μL), while the DNA concentration was determined to be 70.2 ng/μL. Five micrograms of genomic DNA were sheared with a Megaruptor 3 System. The DNA was used to construct a sequencing library using the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences of California, "PacBio", CA, U.S.A.) according to the manufacturer's protocol, and the library was size-selected by using Bluepippin System to obtain a HiFi library with an insert size of 20,000 bp on average. The 248.4 ng of the library was bound on polymerase using the Sequel II Binding kit 2.0 (PacBio), loaded onto Sequel II SMART Cell 8 M (PacBio), and sequenced using the Sequel IIe System (PacBio).

The high-fidelity long reads were obtained by circular consensus sequencing (CCS) analysis from 534 Gbp in total, yielding 120 Gbp of unique molecular sequence, resulting in 31.9 Gbp of HiFi sequences. In this Sequel system, the minimum number of full-length subreads required to generate a HiFi read for a ZMW was equal to 3.

Genome assembly and gene model prediction

The generated PacBio HiFi reads were assembled by the de novo assembler Hifiasm with default parameters[14]. The assembled genome sequence was validated with the Benchmarking Universal Single-Copy Orthologs (BUSCO) software v5.4.4 (https://busco.ezlab.org/) using datasets of actinopterygii_odb10 and vertebrata_odb10[15].

Repeat sequences were identified by using RepeatModeler 2.0.5 and masked by using RepeatMasker 4.1.6 with soft mask option[16]. Protein coding sequences were predicted by BRAKER2 using the de novo RNA-seq transcriptome data reported previously[13,17]. The predicted transcripts were combined and selected by TSEBRA[17,18]. The generated gene model set is named Parran-GS1 and validated with BUSCO software v5.4.4 using datasets of actinopterygii_odb10 and vertebrata_odb10[15]. The gene functions were annotated by Blast2GO tool and EggNOG-mapper implemented in OmicsBox software v3.0 (BioBam Bioinformatics, Valencia, Spain; https://www.biobam.com/omicsbox/)[19–21].

The de novo transcriptome analyses

De novo transcriptomes of *Pempheris nyctereutes* (Pempheridae) and *Glaucosoma buergeri* (Glaucosomatidae) were generated. Total RNA was extracted from fresh muscle tissue using TRIzol (Thermo Fisher Scientific, MA, USA) and purified by on-column DNA digestion using DNase I (QIAGEN) and RNA Clean Concentrator kits (Zymo Research, CA, USA). The mRNA sequencing was performed by Macrogen Japan (Tokyo, Japan). In brief, they constructed sequencing libraries with 1 μg of the RNA samples using TruSeq Stranded mRNA Library Prep kit (Illumina, Foster City, CA, USA) and sequenced using NEXT SEQ 2000 (Illumina) for 151 bp, paired end. The mRNA sequencing of *P. nyctereutes* and *G. buergeri* yielded 47.8 M and 51.7 M reads, 7.23 Gbp and 7.81 Gbp, respectively. The raw reads were quality-filtered by using fastp software version 0.23.2 https://github.com/OpenGene/fastp) with options (-p 30 -u 50) resulting in 47.2 M (6.833 Gbp) and 50.9 M reads (7.32 Gbp) for *P. nyctereutes* and *G. buergeri*, respectively[22,23]. The filtered reads were trimmed first 1 base and the last 2 bases, resulting in each read being 148 bp long. The filtered reads were de novo assembled using Trinity version 2.9.1, and the gene models were predicted by using Transdecoder version 5.5.0[24].

Luciferase gene and HGT analysis

The ostracod luciferase genes were sought from the assembled genome, the gene models, and the raw HiFi reads of *P. ransonneti* using multiple bioinformatic tools with different sensitivities and algorithms to ensure the robustness of our results. Both nucleotide sequences and amino acid sequences of *Cypridina noctiluca* (Genbank Accession Number: BBG57195) were sought against the genome using basic local alignment search tools (BLAST) for high-precision sequence matching[25]. In addition, we performed a search for other ostracod luciferase sequences, *Vargula hilgendorfii* (AAA30332.1) and *Vargula tsujii* (QOS14273.1), using MiniProt[26]. MiniProt was specifically employed to account for potential splice sites and to detect genes that might have been partially fragmented in the assembly or missed during de novo gene prediction by aligning proteins directly to the genomic sequence.

To investigate horizontally transferred genes in *P. ransonneti* from ostracods, we performed a DIAMOND BLAST search of all the genes in the gene model Parran-GS1 against fish and ostracod datasets[27]. DIAMOND was chosen for this genome-wide screening due to its superior computational speed and sensitivity in processing large-scale datasets compared to traditional protein BLAST. The fish dataset includes publicly available proteins of medaka *Oryzias latipes* (GenBank/NCBI RefSeq accession number, GCF_002234675.1), zebrafish *Danio rerio* (GCF_000002035.6), and European sea bass *Dicentrachus labrax* (GCF_905237075.1), and newly generated protein sequences from de novo transcriptome of *P. nyctereutes* and *G. buergeri*. The ostracod dataset includes publicly available proteins of *Darwinula stevensoni* (GCA_905338385.1), *Notodromas monacha* (GCA_905338405.1), *Cyprideis torosa* (GCA_905338395.1), and newly generated protein sequences from the de novo transcriptome of the bioluminescent *V. hilgendorfii*[13].

The potential horizontally transferred genes were further investigated by phylogenetic analysis to determine their evolution. Ortholog groups were predicted using OrthoFinder v2.5.5[28] with the protein sequences of arthropods (*Drosophila melanogaster*, GCF_000001215.4), ostracods (*Notodromas monacha, Darwinula stevensoni*, *Cypridеis torosa*) and teleosts (*Danio rerio*, *Periophthalmus magnuspinnatus* GCF_009829125.3, *Sphaeramia orbicularis* GCF_902148855.1, *Toxotes jaculatrix* GCF_017976425.1, *Epinephelus moara* GCF_006386435.1, *Perca fluviatilis* GCF_010015445.1, *Chelmon rostratus* GCF_017976325.1, *Sparus aurata* GCF_900880675.1) with or without de novo transcriptome data of *V. hilgendorfii*, *P. nyctereutes* and *G. buergeri*. Under the default settings of OrthoFinder 2, gene trees for each orthogroup were inferred using the DendroBLAST algorithm, which converts DIAMOND bit-scores into a distance matrix, followed by tree reconstruction using FastME. The phylogenetic relationship of the species was inferred by STAG in the OrthoFinder2[29], utilizing 2555 single-copy orthologous genes to provide a reliable evolutionary framework for detecting HGT events. This species tree ensured that the baseline phylogeny for the HGT analysis correctly partitioned teleosts and crustaceans into their respective monophyletic clades. The candidate proteins of horizontally transferred genes, which showed higher bit-score in the DIAMOND BLAST search, and luciferases were phylogenetically analyzed with other proteins in the ortholog group. Protein sequences were aligned by MAFFT version 7.490[30] using the BLOSUM62 matrix and trimmed poorly aligned regions using trimAl v.1.5[31] (parameters: -automated1). The maximum likelihood tree was generated by IQ-TREE version 2 using the model chosen according to BIC criteria, and a consensus tree was generated by resampling 1000 replicates of ultrafast bootstrap analysis[32].

Results

High-quality draft genome of *Parapriacanthus ransonneti*

We generated a high-quality draft genome assembly of *P. ransonneti.* The PacBio HiFi library with 20 kbp insert was made from 5 µg of genomic DNA from muscle tissue. We obtained more than 1.8 million HiFi long reads (total 31.9 Gbp) that correspond with 52.1× coverage to the estimated genome size determined to be 613.4 ± 4.7 Mbp by FACS analysis. The Hifiasm analysis resulted in a high-quality assembly with 625 Mbp in total genome, which corresponds to the estimated genome size by FACS analysis. The assembled genome is composed of 424 contigs without gaps. The maximum length of the contigs and N50 were 37.6 Mbp and 13.5 Mbp, respectively (Fig. 2). The detailed statistics of the assembly are shown in Table 1.

The gene model set of *P. ransonneti,* Parran-GS1, was predicted by a combination of using de novo transcriptome and reference-guided prediction. The gene set Parran-GS1 showed a high BUSCO score for both data sets of Actinopterygii (96.3% complete benchmarking universal single-copy orthologs, also known as BUSCOs, were recovered) and vertebrata (96.5% complete BUSCOs were recovered) (Table 1).

Phylogenetic relationship of *P. ransonneti*

Phylogenetic relationship inferred by STAG in Orthofinder2 with 2555 gene trees recovered the monophyly of both Actinopterygii and Arthropoda (Fig. 3A). *Parapriacanthus ransonneti* were the most closely related to *Pempheris nyctereutes* to form the monophyletic family Pempheridae. *Glaucosoma buergeri* was the most closely related species to the Pempheridae.

Absence of ostracod luciferase in the *P. ransonneti* genome

To test the hypothesis that the luciferase genes of ostracods are absent in the genome of *P. ransonneti*, we performed a homology search for the luciferases against the generated data set derived from the genome assembly of *P. ransonneti.* The nucleotide and amino acid sequences of luciferases of *C. noctiluca*, *V. hilgendorfii,* and *V. tsujii* were used as query sequences, although our previous study detected peptide sequences of the fish luciferase identical to that of *C. noctiluca* and differing from that of *V. hilgendorfii.*

First, we performed BLASTp search against the protein dataset translated from the Parran-GS1 gene model. No matches to luciferases from *C. noctiluca* or other ostracods were found. The top hit for the luciferase of *C. noctiluca* was Pr_v1_g20265.t2 (bitscore 75.1 and e-value 9e−14), showing 12% similarity (BLOSUM 62) and 5.5% identity (Fig. 3B). This protein does not contain the sequences matching to the detected peptides from mass spectrometry analysis of the luciferase in light organs of *P. ransonneti*[13]. The most similar gene of Pr_v1_g20265.t2 in the zebrafish *D. rerio* is IgG Fc-binding protein X3 (XP_017211776.2, bitscore 1334 and e-value 0.0). The luciferase activity has never been reported from IgGFc-binding protein X3. The phylogenetic tree of these genes showed a consistent relationship to the species tree, suggesting that these genes are not horizontally transferred from *C. noctiluca* to *P. ransonneti* (Fig. 3C). These results show that Pr_v1_g20265.t2 is not a gene encoding the luciferase of *C. noctiluca*.

We sought ostracod luciferase genes against the genome assembly of *P. ransonneti*. To validate the gene models and search methods, genes such as sidky (SH3), pannexin 2 (Panx2), sorting nexin (SNX33), FIC domain (ficd), histone H3, and glycosyltransferase (GLYT) were sought by tBLASTn search and detected in the genome of *P. ransonneti*. The tBLASTn search did not detect any luciferase gene of *C. noctiluca* and *V. tsujii* in the genome of *P. ransonneti*. The tBLASTn search for luciferase of *V. hilgendorfii* found one hit on the contig ptg000015 (position 5469538 to 5469299) with the e-value 6e−5 and bitscore 50.4. The most similar sequence to the luciferase of *V. hilgendorfii* in the gene model of *P. ransonneti* is Pr_v1_g7343.t1, showing 8.8% similarity (BLOSUM 62) and 3.7% identity. The genomic position of this gene is 5457978–5480911 on contig ptg000015. The most similar gene of this gene in the zebrafish *D. rerio* encodes IgGFc-binding protein (XP_021322199, bitscore = 1575 and e-value = 0.0). In addition, we applied MiniProt, a protein-to-genome alignment algorithm that considers exon–intron structure, to search the *P. ransonneti* genome using ostracod luciferase queries. However, MiniProt also failed to detect any luciferase genes from ostracods.

We also sought the ostracod luciferase genes from pre-assembled HiFi long reads. Three ostracod luciferases from *C. noctiluca, V. hilgendorfii,* and *V. tsujii* were not mapped on any HiFi reads by MiniProt. The MiniProt

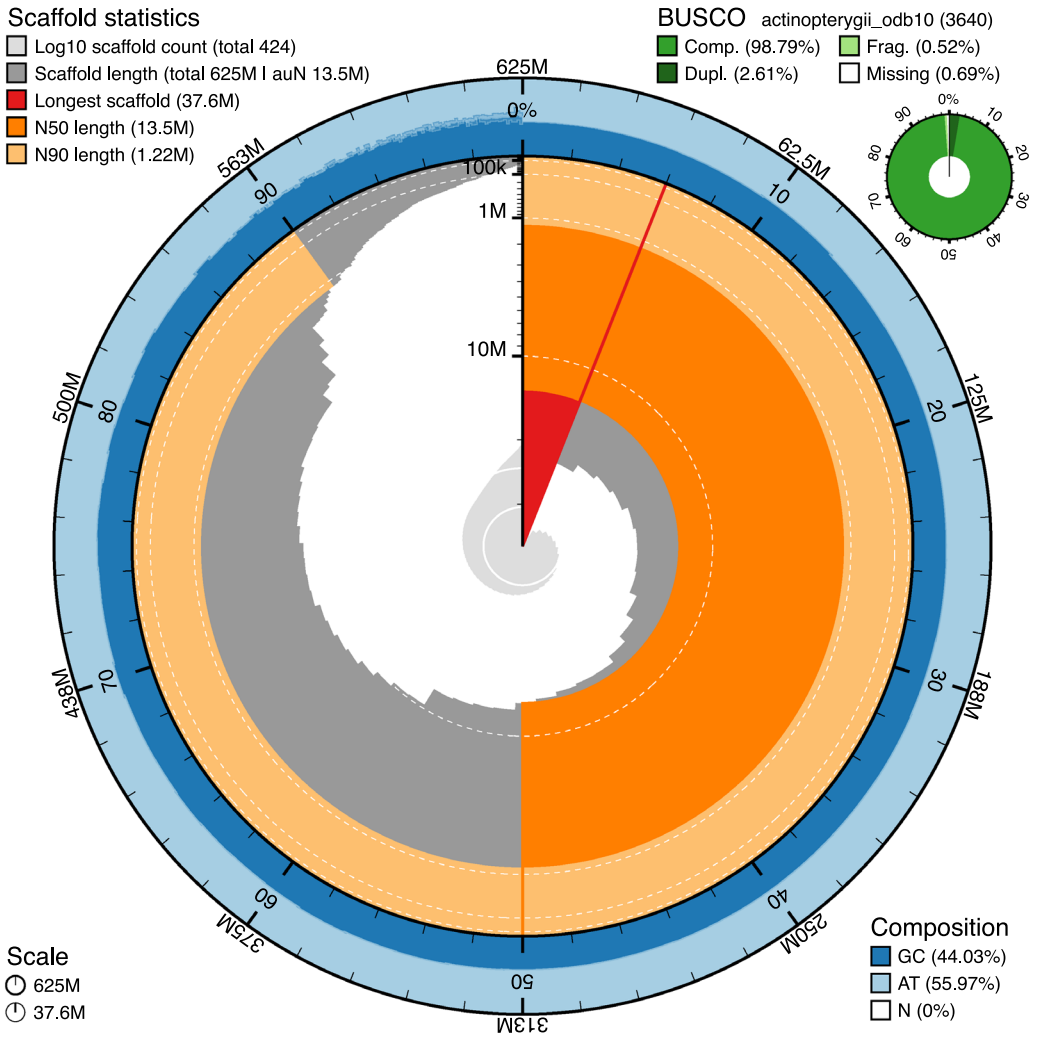

Fig. 2. The *Parapriacanthus ransonneti* genome assembly statistics. The snail plot summarizes the contiguity, completeness, and composition of the *P. ransonneti* genome assembly. The main grey spiral represents the cumulative scaffold length, sorted by size, with the total assembly length reaching 625 Mbp. The scaffold N50 is displayed in dark orange (13.5 Mbp), N90 in light orange (1.22 Mbp), and GC content in dark blue (44.03%). The circular plot at the top right indicates high genome completeness with a BUSCO score of 98.79% (actinopterygii_odb10).

search was also performed for IgGFc-binding protein homologs in *P. ransonneti*, Pr_v1_g20265.t2 and Pr_v1_g7343.t1, as positive control, which resulted in three and two HiFi read hits, respectively. The tBLASTn search for luciferase of *C. noctiluca* against the HiFi reads resulted in one hit on one read (CCS ID = 94897028; bitscore = 37 and e-value 6e−4). This read showed a higher bitscore (480) and lower e-value (0.0) when the Pr_v1_g7343.t1 gene is used as a query, which suggests that the corresponding region encodes fish IgGFc-binding protein rather than luciferase of *C. noctiluca*.

None of the searches described above detected ostracod luciferase genes in the gene set, whole genome, and preassembled reads. Although the two sequences—Pr_v1_g20265.t2 and Pr_v1_g7343.t1—were found as top hits in the tBLASTn search, both encode IgGFc-binding protein rather than luciferases.

To further investigate the prevalence of ostracod-type luciferase sequences among fishes, we performed a BLASTp search against the NCBI non-redundant (nr) protein database, specifically targeting the Teleostei (Taxid: 32443). Consistent with our findings in *P. ransonneti*, the top hits for all teleost species were IgG Fc-binding proteins (e.g., *Oreochromis niloticus* IgG Fc-binding protein-like, CAI5677446.1). No sequences matching functional cypridinid luciferases were identified across the teleost lineage. Furthermore, our functional annotation using the OmicsBox suite, which integrates Blast2GO and EggNOG-mapper, did not identify any proteins with the specific domain architecture characteristic of cypridinid luciferases among the 28,610 predicted proteins. While proteins containing the von Willebrand factor type D (VWD) domain were identified, they were consistently annotated as fish IgG Fc-binding proteins rather than functional luciferases. These genome-

Statistic	Values
Assembly size	625,342,690 bp
Max length	37,616,461 bp
N50	13,506,590 bp
Sum of gap	0 bp
BUSCO results—genome	
vertebrata_odb10	C: 98.1% [S: 94.4%, D: 3.7%], F: 0.7%, M: 1.2%, n: 3354
actinopterygii_odb10	C: 98.8% [S: 96.2%, D: 2.6%], F: 0.5%, M: 0.7%, n: 3640
Genes in Parran-GS1	28,610
Sum length of genes	44,681,749 bp
Average length of genes	1561.80 bp
Max length of genes	75,504 bp
N50 of genes	2280 bp
Number of transcripts	37,210
BUSCO results—protein	
vertebrata_odb10	C: 96.5% [S: 72.3%, D: 24.2%], F: 1.6%, M: 1.9%, n: 3354
actinopterygii_odb10	C: 96.3% [S: 72.7%, D: 23.6%], F: 1.1%, M: 2.6%, n: 3640

Table 1. General statistics of assembled genome of *Parapriacanthus ransonneti. C* complete, *S* single-copy, *D* duplicated, *F* fragmented, *M* missing, *n* number of gene searched.

wide and cross-species analyses provide robust evidence that *P. ransonneti* lacks the genetic capacity to produce luciferase endogenously, supporting the hypothesis of protein sequestration from its prey.

Investigation of horizontally transferred gene from ostracods to *P. ransonneti*

Bioluminescence reaction in *P. ransonneti* involves luciferin and luciferase. Although no genes related to regulating sequestered luciferin have yet been identified, the lateral acquisition of such genes may confer an adaptive advantage for utilizing bioluminescence. We therefore hypothesized that *P. ransonneti* might have acquired additional genes via horizontal gene transfer (HGT) from luminous ostracods, beyond the luciferase genes.

To test this hypothesis, we searched for candidate horizontally transferred genes (xenologs) by performing DIAMOND BLAST comparisons against RefSeq databases for both fish and ostracods, using all predicted proteins from the Parran-GS1 gene set. For each predicted protein, bitscores were compared between the two reference sets, revealing 20 proteins with higher similarity to ostracod sequences than to fish sequences.

Phylogenetic analyses were performed on these 20 candidate proteins to evaluate their evolutionary relationships. However, no evidence supporting horizontal gene transfer (HGT) from the ostracods to *P. ransonneti* was observed. Specifically, these 20 genes showed a sister relationship to the other fish genes but were not nested in the ostracods clades. These analyses did not indicate that genes from ostracods had been transferred to the genome of *P. ransonneti* or other members of the Pempheridae family. In conclusion, no evidence of horizontal gene transfer from the ostracods was detected in the genome of *P. ransonneti.*

Discussion

To investigate the presence of the ostracod luciferase gene in the genome of *Parapriacanthus ransonneti,* we performed genome sequencing and horizontal gene transfer analysis. We provided a high-quality draft genome of *Parapriacanthus ransonneti* by using PacBio HiFi reads. The assembled genome size (625 Mbp) was in good agreement with the estimated genome size (613 Mbp) measured by FACS analysis. The 50 longest contigs account for 90% of the total genome without gaps. The high BUSCO score suggests that our draft genome is adequate to examine the presence or absence of the luciferase gene. The model of the gene set, Parran-GS1, generated by a combination of de novo transcriptome and in silico prediction, showed a high BUSCO score, suggesting that it recovered most of the genes of *P. ransonneti.*

Parapriacanthus ransonneti possesses and uses the luciferase; the amino acid sequence is identical to that of a luminous ostracod *C. noctiluca*. The luciferase activity in the light organ was lost in a year when the fish was maintained without luminous ostracods as food. Feeding closely related luminous ostracod *V. hilgendorfii* replaces the peptide sequences of the luciferase of *C. noctiluca* with those of *V. hilgendorfii.* This replacement is best explained by sequestration of ingested proteins (kleptoprotein). An alternative hypothesis, however, is that *P. ransonneti* possesses endogenous luciferase genes corresponding to both ostracod species and differentially expresses them after ingesting each prey. To rule out this alternative hypothesis, previously we attempted to detect luciferase genes and their transcripts from *P. ransonneti,* but all the attempts resulted in negative: luciferase gene was not found by BLAST search against transcriptome of *P. ransonneti* generated by de novo mRNA-seq; in vitro expression assay using rabbit reticulocyte and wheat germ extract with total RNA of *P. ransonneti* gave no luciferase activity; genomic PCR with five different primer sets for the luciferase gene did not amplify luciferase gene fragments[13]. These results support a hypothesis of the absence of the ostracod luciferase gene in the genome and transcriptome of *P. ransonneti*.

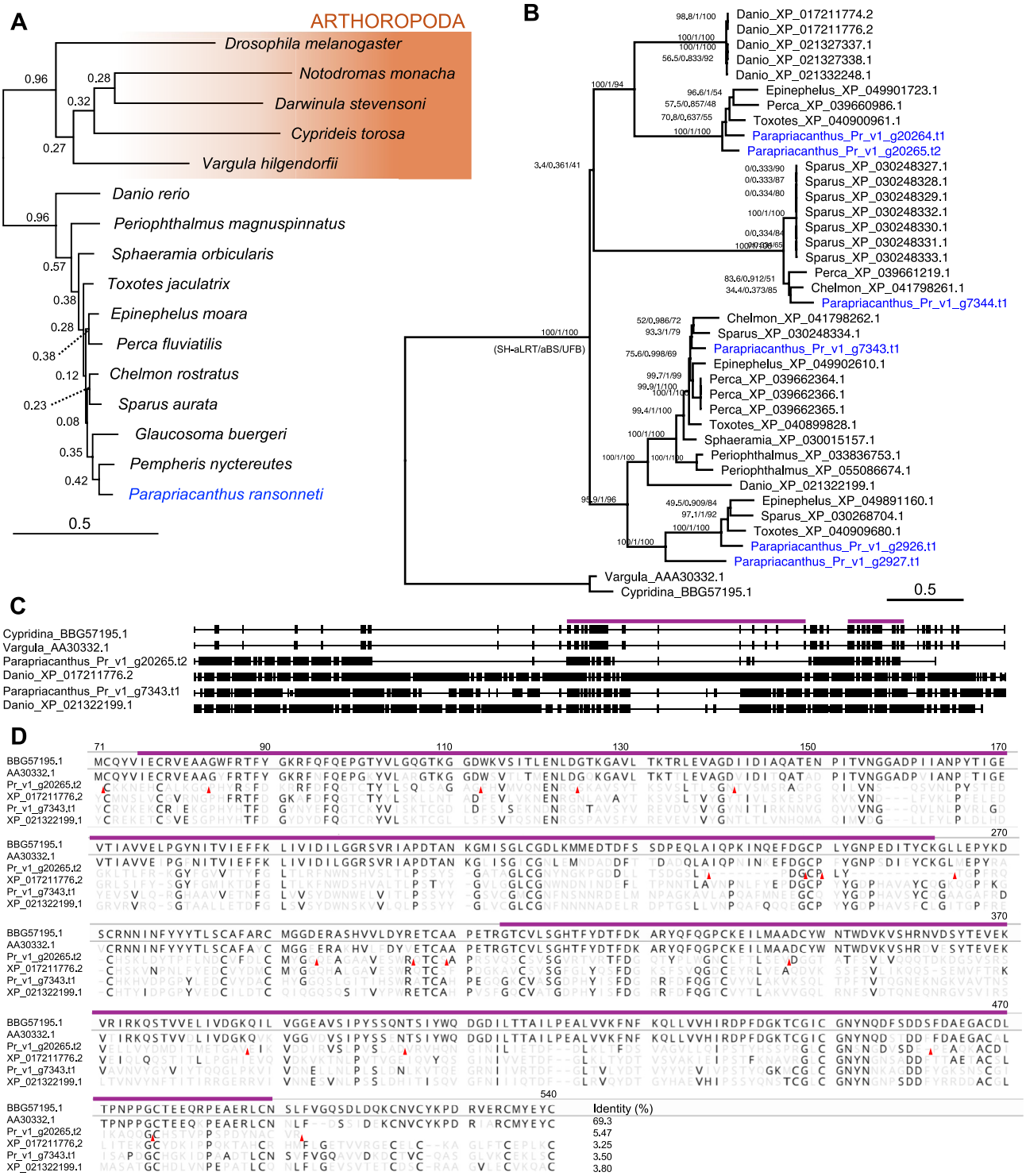

Fig. 3. Comparative genome analysis for ostracod luciferase. (**A**) Phylogenetic relationships among the species used in this study to detect horizontal gene transfer. The species tree was inferred using STAG based on 2555 single-copy ortholog gene trees, reconstructed via DendroBLAST and FastME under OrthoFinder. The node support values represent the proportion of gene trees that support each bipartition in the species tree. (**B**) Maximum Likelihood (ML) tree of Cypridinid luciferases and their homologs. Numbers in parentheses at the nodes are SH-aLRT support (SH-aLRT), aBayes support (aBS), and ultrafast bootstrap support (UFB). GenBank accession numbers are listed next to each genus name. (**C**,**D**) Overview (**C**) and close-up view (**D**) of the multiple sequence alignment of ostracod luciferases and their homologs in fishes. Two von Willebrand factor type D domains (indicated by purple lines) are conserved in the ostracod luciferases. The most similar protein in *Parapriacanthus ransonneti* to the ostracod luciferase contains gaps (red triangles) ranging from 1 to 806 amino acids, with an average gap size of 58.9 amino acids (standard deviation: 187).

Although the earlier experiments strongly suggested that ostracod luciferase genes are absent in the genome of *P. ransonneti,* they could not fully exclude the possibility that the negative results arose from technical limitations. The whole genome sequencing of *P. ransonneti* was necessary to determine which hypothesis is correct.

In this study, we performed tBLASTn and BLASTp searches for the ostracod luciferase gene from the generated gene models, the assembled genome, and pre-assembled HiFi long reads, but did not detect the luciferase gene. We also sought the luciferase genes from the genome assembly and pre-assembled HiFi long reads by MiniProt but did not find them. We obtained 31.9 Gbp of the HiFi reads corresponding to 52.1 times coverage of the genome of *P. ransonneti.* Critically, the absence of hits even in the raw HiFi reads—which bypass potential assembly artefacts—provides compelling evidence that ostracod luciferase genes are truly absent from the genome of *P. ransonneti*.

Taken together, we conclude that the ostracod luciferase gene is absent in the genome of *P. ransonneti*, supporting the hypothesis that *P. ransonneti* cannot produce luciferase by itself but instead sequesters and uses the luciferase from ostracod prey for its bioluminescence.

The phylogenetic analysis by STAG using 2555 single-copy orthologous genes suggested that *P. ransonneti* forms a clade with *Pempheris nyctereutes,* which is sister to *Glaucosoma buergeri.* The datasets for *P. nyctereutes and G. buergeri* were derived from transcriptomes, which do not include data from genome sequencing data and are likely to contain isoforms, resulting in a reduced number of identifiable single-copy orthologs. The wide taxon sampling from Arthropoda and Vertebrata causes low numbers of the single-copy orthologous genes, which perhaps affect the tree topology of phylogenetic analysis. However, the phylogenetic analysis using 7881 single-copy orthologous genes excluding those of arthropod genomes, and those from transcriptomes of *P. nyctereutes* and *G. buergeri,* showed consistent tree topology, where *P. ransonneti* was sister to the clade including *Chelmon rostratus* and *Sparus aurata.* This robust separation of the fish lineage from the crustacean lineage confirms the accuracy of our phylogenetic framework, ensuring that subsequent orthology-based analyses and the detection of xenologs were performed reliably and without taxonomic ambiguity.

Using this gene set, we sought horizontally transferred genes (xenologs) from ostracod to *P. ransonneti.* The potential xenologs in *P. ransonneti* from ostracods were listed by comparing the bitscore of BLASTp search analysis. None of those potential xenologs showed tree topology, which suggests the horizontal gene transfer from crustacean/ostracods to Pempheridae fishes, by phylogenetic analysis.

Bioluminescence reaction requires not only luciferase but also luciferin as a substrate (Fig. 1C). *Cypridina* luciferin—also known as cypridinid luciferin, *Vargula* luciferin, and vargulin—is involved in the light-emitting reaction in the luminous ostracods, such as *Vargula hilgendorfii* and *Cypridina noctiluca,* and the fish *P. ransonneti.* In addition to luciferin, the luminous ostracods contain sulfated luciferin, which does not have luminescent activity but is more stable than luciferin, the active form[33]. The stabilized luciferin might be a storage form and activated on demand by luciferin sulfotransferase[34]. *Vargula hilgendorfii* can de novo synthesize its luciferin, whereas *P. ransonneti* must acquire it through diet[13,35]. The recruitment of sulfotransferase for luciferin has been repeatedly evolved among multiple lineages of luminous organisms[34,36,37]. We hypothesized that *P. ransonneti* may have acquired genes related to bioluminescence from the luminous ostracods for efficient regulation of bioluminescence. To test this hypothesis, we performed an analysis to detect xenologs in the genome of *P. ransonneti* from ostracods. However, our analysis did not detect xenologs, including sulfotransferase genes, transferred from ostracods in the *P. ransonneti* genome. This suggests that *P. ransonneti* evolved to have its own genetic bases for regulating bioluminescence with sequestered luciferase.

The sequestration and use of prey protein, known as kleptoprotein, is reported only from *Parapriacanthus ransonneti.* Our high-quality draft genome provided evidence that the ostracod luciferase gene is absent in the genome of *P. ransonneti.* Our findings revealed that the acquisition of the function of the prey organism can be achieved without horizontal gene transfer. The molecular mechanisms underlying protein sequestration are not elucidated yet. The genome sequence of *P. ransonneti* provides a fundamental source to understand the evolution of kleptoprotein bioluminescence.

Data availability

The raw read data and assembled genomes of *P. ransonneti* are available in the DDBJ/ENA/GenBank database (BioProject PRJDB35272). The raw read data for the mRNA-seq of *P. nyctereutes* and *G. buergeri* are available in the DDBJ/ENA/GenBank database (BioProject PRJDB38047).

Received: 19 December 2025; Accepted: 9 March 2026
Published online: 01 April 2026

References

1. Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids; A structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).
2. Morgan, Sturtevant, Muller & Bridges. Mechanism of Mendelian heredity. *Trans. Am. Microsc. Soc.* **34**, 293 (1915).
3. Visscher, P. M., Hill, W. G. & Wray, N. R. Heritability in the genomics era—Concepts and misconceptions. *Nat. Rev. Genet.* **9**, 255–266 (2008).
4. Christmas, M. J. et al. Evolutionary constraint and innovation across hundreds of placental mammals. *Science* **380**, eabn3943 (2023).
5. Moran, N. A. Symbiosis. *Curr. Biol.* **16**, R866–R871 (2006).
6. McFall-Ngai, M. J. The importance of microbes in animal development: Lessons from the squid-vibrio symbiosis. *Annu. Rev. Microbiol.* **68**, 177–194 (2014).
7. Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y. & Ishikawa, H. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* **407**, 81–86 (2000).

8. Goodheart, J. A. & Bely, A. E. Sequestration of nematocysts by divergent cnidarian predators: Mechanism, function, and evolution. *Invertebr. Biol.* **136**, 75–91 (2017).
9. Goodheart, J. A., Barone, V. & Lyons, D. C. Movement and storage of nematocysts across development in the nudibranch *Berghia stephanieae* (Valdés, 2005). *Front. Zool.* **19**, 16 (2022).
10. Goodheart, J. A. et al. A chromosome-level genome for the nudibranch gastropod *Berghia stephanieae* helps parse clade-specific gene expression in novel and conserved phenotypes. *BMC Biol.* **22**, 9 (2024).
11. Trench, R. K., Greene, R. W. & Bystrom, B. G. Chloroplasts as functional organelles in animal tissues. *J. Cell Biol.* **42**, 404–417 (1969).
12. Maeda, T. et al. Chloroplast acquisition without the gene transfer in kleptoplastic sea slugs, *Plakobranchus ocellatus*. *Elife* **10**, e60176 (2021).
13. Bessho-Uehara, M. et al. Kleptoprotein bioluminescence: *Parapriacanthus* fish obtain luciferase from ostracod prey. *Sci. Adv.* **6**, eaax4942 (2020).
14. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).
15. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).
16. Flynn, J. M. et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9451–9457 (2020).
17. Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom. Bioinform.* **3**, lqaa108 (2021).
18. Gabriel, L., Hoff, K. J., Brůna, T., Borodovsky, M. & Stanke, M. TSEBRA: Transcript selector for BRAKER. *BMC Bioinform.* https://doi.org/10.1101/2021.06.07.447316 (2021).
19. Götz, S. et al. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **36**, 3420–3435 (2008).
20. Huerta-Cepas, J. et al. eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).
21. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. EggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).
22. Chen, S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *Imeta* **2**, e107 (2023).
23. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
24. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
25. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
26. Li, H. Protein-to-genome alignment with miniprot. *Bioinformatics* https://doi.org/10.1093/bioinformatics/btad014 (2023).
27. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
28. Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
29. Emms, D. M. & Kelly, S. STAG: Species tree inference from all genes. *bioRxiv* https://doi.org/10.1101/267914 (2018).
30. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
31. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
32. Minh, B. Q. et al. Corrigendum to: IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 2461 (2020).
33. Nakamura, M., Suzuki, T., Ishizaka, N., Sato, J.-I. & Inouye, S. Identification of 3-enol sulfate of *Cypridina* luciferin, *Cypridina* luciferyl sulfate, in the sea-firefly *Cypridina (Vargula) hilgendorfii*. *Tetrahedron* **70**, 2161–2168 (2014).
34. Lau, E. S. et al. Similar enzymatic functions in distinct bioluminescence systems: Evolutionary recruitment of sulfotransferases in ostracod light organs. *Biol. Lett.* **20**, 20230585 (2024).
35. Kato, S.-I., Oba, Y., Ojika, M. & Inouye, S. Identification of the biosynthetic units of *Cypridina* luciferin in *Cypridina (Vargula) hilgendorfii* by LC/ESI-TOF-MS. *Tetrahedron* **60**, 11427–11434 (2004).
36. Lau, E. S. & Oakley, T. H. Multi-level convergence of complex traits and the evolution of bioluminescence. *Biol. Rev. Camb. Philos. Soc.* **96**, 673–691 (2021).
37. Tzertzinis, G. et al. Coelenterazine sulfotransferase from *Renilla muelleri*. *PLoS ONE* **17**, e0276315 (2022).

Acknowledgements

We thank Dr. Akinori Teramura in Tokai University, Dr. Hidetoshi Wada in Kanagawa Prefectural Museum, and Jian-Fu Huang for supporting the sample collection of specimens of *P. nyctereutes* and *G. buergeri*, and to Dr. Keiichi Sato in Okinawa Churaumi Aquarium for providing specimens of *P. ransonneti*. We also thank Dr. Kanako Bessho-Uehara at Tohoku University and Dr. Tsutomu Tsuchida at Toyama University for critical advice to improve analyses and the manuscript. We acknowledge NIBB for technical support as a part of the Trans-Scale Biology program. Computations were partially performed on the NIG supercomputer at ROIS National Institute of Genetics, NIBB supercomputer at the Data Integration and Analysis Facility, National Institute for Basic Biology (Project: 22-NIBB-103), and RCCS supercomputer at Research Center for Computational Science, Okazaki, Japan (Project: 24-IMS-C380).

Author contributions

M.B.-U. conceived and designed this study and wrote the manuscript. M.B.-U., K.K., and S.M. collected and prepared the tissue samples for sequencing. M.B.-U. and K.K. collected transcriptome data. M.B.-U., K.Y., T.M. and S.S. performed genome size estimation, genome sequencing, bioinformatics analysis. All the authors have read and approved the final manuscript.

Funding

This study was supported in part by JST FOREST Program (Grant Number: JPMJFR214D), JSPS KAKENHI (Grant Number: 20K22627, 21K15144, 21K06313), and NIBB Collaborative Research Program (Grant Number: 23NIBB103, 24NIBB102).

Declarations

Competing interests
The authors declare no competing interests.

Ethical approval
All animal experiments were approved by the Institutional Animal Experiment Committee of Nagoya University (Approval Nos. S210622, S220019, and S230010). All methods were performed in accordance with the university's regulations and the ARRIVE guidelines. Detailed descriptions of the ethical procedures and animal handling are provided in the "Materials and methods" section.

Additional information
Correspondence and requests for materials should be addressed to M.B.-U.

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.