Zoological Studies **62:**25 (2023)
doi:10.6620/ZS.2023.62-25

Zoological Studies

Open Access

Long-Read Genome Sequencing of *Abscondita cerata* (Coleoptera: Lampyridae), the Endemic Firefly of Taiwan

Tzi-Yuan Wang[1], King-Siang Goh[2], Liang-Jong Wang[3], Li-Ling Wu[4], Feng-Yu Wang[5,*], and Yu-Wei Wu[6,7,8,*]

[1]*Biodiversity Research Center, Academia Sinica, Taipei 115, Taiwan. E-mail: tziyuan@gmail.com (TY Wang)*
[2]*Genomics Research Center, Academia Sinica, Taipei 115, Taiwan. E-mail: gohks0505@gmail.com (Goh)*
[3]*Forest Protection Division, Taiwan Forestry Research Institute, Taipei 100, Taiwan. E-mail: josephwang23@tfri.gov.tw (LJ Wang)*
[4]*Department and Institute of Physiology, College of Medicine, National Yang Ming Chiao Tung University, Taipei 112, Taiwan. E-mail: wuliling@nycu.edu.tw (LL Wu)*
[5]*Taiwan Ocean Research Institute, National Applied Research Laboratories, Kaohsiung 852, Taiwan. *Correspondence: E-mail: hitachiwang@gmail.com (FY Wang)*
[6]*Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei 106, Taiwan. *Correspondence: Email: yuwei.wu@tmu.edu.tw (YW Wu)*
[7]*Clinical Big Data Research Center, Taipei Medical University Hospital, Taipei 235, Taiwan*
[8]*TMU Research Center for Digestive Medicine, Taipei Medical University Taipei 110, Taiwan*

Received 27 August 2022 / Accepted 17 February 2023 / Published 26 May 2023
Communicated by Chih-Ming Hung

Abscondita cerata is the most abundant and widely distributed endemic firefly species in Taiwan and is considered a key environmental and ecological indicator organism. In this study, we report the first long-read genome sequencing of *Abs. cerata* sequenced by Nanopore technology. The draft genome size, 967 Mb, was measured through a hybrid approach that consisted of assembling using 11.25-Gb Nanopore long reads and polishing using 9.47-Gb BGI PE100 short reads. The drafted genome was assembled into 4,855 contigs, with the N50 reaching 325.269 kb length. The assembled genome was predicted to possess 55,206 protein-coding genes, of which 20,862 (37.78%) were functionally annotated with public databases. 47.11% of the genome sequences consisted of repeat elements; among them DNA transposons accounted for the largest proportion (26.79%). A BUSCO (Benchmarking Universal Single Copy Orthologs) evaluation demonstrated that the genome and gene completeness were 84.8% and 79%, respectively. The phylogeny constructed using 1,792 single copy genes was consistent with previous studies. The comparative transcriptome between adult male head and lantern tissues revealed (1) the vision of *Abs. cerata* is primarily UV-sensitive to environmental twilight, which determines when it begins its nocturnal activity, (2) the major expressed *OR56d* receptor may be correlated to suitable humidity sensing, and (3) *Luc1*-type luciferase is responsible for *Abs. cerata*'s luminescent spectrum.

Key words: Nanopore, firefly, Lampyridae, *Abscondita*, Genome, Vision, Olfaction, Luciferase.

BACKGROUND

Abscondita cerata (Fig. 1) is an abundant *Abscondita* species (Ballantyne et al. 2013 2019) that ranges widely from sea level to 1,500 m above sea level and is endemic to the country side of Taiwan island (Ohba and Yang 2003). Its mating season is mainly from March to May (Wu et al. 2010). *Abs. cerata* is most active during the nighttime around 18:30–23:00 (Goh et al. 2022a; Ohba and Yang 2003). Several cohabitated species also can be found during its mating season, such as the genus *Luciola*, *Aquatica*,

Citation: Wang TY, Goh KS, Wang LJ, Lee CM, Wu LL, Wang FY, Wu YW. 2023. Long-Read Genome Sequencing of *Abscondita cerata* (Coleoptera: Lampyridae), the endemic firefly of Taiwan. Zool Stud **62:**25. doi:10.6620/ZS.2023.62-25.

Abscondita, *Pyrocoelia*, *Curtos* (Goh et al. 2022ab; Jeng et al. 1998 1999; Ohba and Yang 2003). Among the cohabitated species in Taiwan, male *Abs. cerata* produce the brightest flashes (luminescence emissions λ_{max} = 563.6 ± 0.3 nm), measuring up to 14 lux (Goh et al. 2022b). *Abs. cerata* starts its nocturnal activity when the environmental light intensity is lower than 6.49 lux (Goh et al. 2022b). Comparative morphological characters provided evidence to support the transfer of six *Luciola* species (including *Abs. cerata*) into *Abscondita* (Ballantyne et al. 2013). Two new species of *Abscondita* and *Luc. pallescens* were also included in a later study (Ballantyne et al. 2019). In addition, the mitochondrial cytochrome *c* subunit I (*COI*) barcode with 3.27%–12.3% "intraspecific" variations revealed the uniqueness of Taiwanese fireflies, such as *Abs. chinensis*, *Aquatica ficta*, *Luc. curtithorax*, *Luc. filiformis* and *Curtos costipennis* (Goh et al. 2022b). Thus, more genome comparison could help to solve the phylogenetic placements of these cryptic Asian fireflies.

A total evidence phylogenetic approach revealed the evolution of adult luminescence in fireflies (Martin et al. 2017). There are two molecular bases for firefly luminescence communication: one gene family (opsins) controls the detection of the signal, and another gene family (luciferase) controls the production of the signal (Martin et al. 2015). The major insect opsin class (SW) that typically confers sensitivity to "blue-light" was lost before the origin of modern beetles, whereas duplications within the *UV* opsin class have likely overcome the loss of blue sensitivity (Sharkey et al. 2017). Thus, the color vision of fireflies are mainly controlled by two family genes, the 480–600 nm long-wavelength-sensitive (*LWS*) opsin and the 300–400 nm ultraviolet-sensitive (*UVS*) opsin (Owens and Lewis 2018). Previous studies also supported the idea that short- and mid-wavelength artificial light may influence the flash signal of fireflies, such as *Aq. ficta*, which can sense these light wavelengths (Owens et al. 2018). On the other hand, two paralogous luciferases (Luc1 and Luc2) were identified in fireflies, in which Luc1-type luciferase was responsible for producing the "yellowish" luminescence of adult firefly lanterns while Luc2-type luciferase was used to produce the distinct dim, greenish glow of eggs and the whole pupa (Bessho-Uehara et al. 2017; Oba et al. 2013).

High-throughput genome sequencing is an important milestone for enhancing the quality and speed of traditional genome research, in which both next- and third-generation sequencing platforms contribute to accelerating the study of draft genome sequences of non-model organisms. This study demonstrates a hybrid genome assembly of the firefly *Abs. cerata*, which is an endemic species of *Abscondita* in Lampyridae, by using low coverage of long-read sequences followed by error correction using reasonable coverage of short-read sequences. The phylogenetic and comparative genomic analyses provided insights into the evolution of luciferase, opsin genes and olfactory receptor (*OR*) genes in fireflies. The draft genome also serves as a valuable resource for the molecular basis of firefly luminescence and ecology conservation.

Fig. 1. *Abscondita cerata* male photographed by Shih-Chieh Huang in Chienshih Township, Hsinchu County, Taiwan.

MATERIALS AND METHODS

Specimen

Seventeen adult male specimens were collected from the habitat in Nankang, Taipei City, Taiwan (25°01'40.4"N 121°38'02.6"E) in April 2019. After they were collected from the field, each living specimen was transferred and stored in a 50 mL sampling tube with moist paper tissues in the laboratory, under a 12L:12D photoperiod at 25°C. To obtain the head and lantern tissues used for extracting RNA and genomic DNA, each specimen was sacrificed as previously described (Goh and Li 2011). The isolated head or lantern was immediately preserved via Invitrogen™ RNAlater™ Stabilization Solution (Cat No. AM7021) in a -80°C freezer for further RNA extraction. The remaining thorax and body tissues were freshly preserved in a -20°C freezer for further genomic DNA extraction.

Genome size estimated by propidium iodide (PI) DNA staining with flow-cytometer

The fresh head tissues were dissected in 2 mL filtered Galbraith buffer (dissolve 4.26 g $MgCl_2$, 8.84 g sodium citrate, 4.2 g 3-[N-morpholino] propane sulfonic acid ("MOPS"), 1 mL Triton X-100, and 1 mg boiled ribonuclease A into 1 L of ddH_2O with a final pH of 7.2), and the debris was filtered via 5-mL round bottom tubes (Falcon Cat# 352235). The filtered cell solution was centrifuged at a low rate of 1500–3000 rpm at 4°C for 10 minutes. After removing the supernatant, remaining cells were re-suspended in 2 mL of 2% paraformaldehyde for 10–30 minutes then centrifuged again and resuspended with 1 mL of 1X PBS (Phosphate Buffered Saline) for PI staining (Orgogozo and Rockman 2011). 10,000 events of stained cells were analyzed on BD FACS Verse flow cytometer. During analysis, aggregates were removed by applying a gate for doublet elimination. The percentage of cells in G0/G1, S, G2/M and Sub G1 phases of the cell cycle was determined using BD FACSDiva 8.0.2 software (Becton, Dickinson and Company BD Biosciences, USA). The genome size estimation was conducted following the protocol as previously described (Orgogozo and Rockman 2011) and normalized using a *Danio rerio* AB strain (genome size = 1.412 Gb).

Whole-genome sequencing

The genomic DNA of thorax tissue was extracted via the ZR Tissue & Insect DNA MicroPrep™ kit (D6015) following the supplier's instructions. The extracted genomic DNA was checked in 0.8–1% agarose gel and concentrated by the size selection of KAPA Pure Beads (Cat# KR1245). For long-read sequencing, the concentrated genomic DNA (1–3 µg) was end-repaired and ligated with Nanopore sequencing adaptors (AMX in Nanopore Ligation Sequencing Kit, SQK-LSK109) via a KAPA Hyper Prep Kit (Cat#KR0961, Kapa Biosystems, Wilmington, MA), following the manufacturer's instructions. The genomic DNA library was premixed with the LB and SQB of the Nanopore Ligation Sequencing Kit (SQK-LSK109), loaded in a Flow Cell (R9.4.1; FLO-MIN106), and sequenced by MinION devices for 48–72 h in laboratory. The BGISEQ-500RS short reads data were referred to in a previous study (Wang et al. 2021) and can be accessed at NCBI SRA (accession no. SRR13626958).

Transcriptomic sequencing

The RNA of each tissue was grinded in Invitrogen™ TRIzol™ Reagents (Cat. No. 15596026) and then extracted using a Qiagen RNeasy Mini Kit (Cat. No. 74104) with in-silico DNase I incubation, following the manufacturer's instructions. The library of purified mRNA was constructed using a TruSeq Stranded mRNA Library Prep Kit (Illumina, San Diego, CA, USA), following the manufacturer's instructions. The constructed libraries were sequenced on an Illumina NovaSeq 6000 platform with 150 bp paired-end reads generated by Genomics, BioSci & Tech Co., New Taipei City, Taiwan.

Genome Assembly and Annotation

Nanopore reads were trimmed using Porechop v0.2.4 (Wick et al. 2017) and assembled using canu v2.2 (Koren et al. 2017) (parameter: -nanopore-raw). The assembled contigs were corrected by the following procedure. The assembly was first corrected by Racon v1.4.3 (Vaser et al. 2017) for four rounds (parameters: -m 8 -x -6 -g -8 -w 500; alignment was made using minimap2 v2.22-r1101 (Li 2018)) followed by one round of Medaka v1.4.4 correction (using medaka_consensus) (https://github.com/nanoporetech/medaka). The BGI paired-end reads (Wang et al. 2021) were trimmed using SOAPnuke v2.0.7 (Chen et al. 2018) with parameters "-l 10 -q 0.5 -n 0.1 -M 2 --adaMR 0.5 -f AAGTCGGAGGCCAAGCGGTCTTAGGAAGACAA -r AAGTCGGATCGTAGCCATGTCGTTCTGTGAGCCAAGGAGTTG" and mapped against the racon/medaka-corrected contigs using bowtie2 v2.2.3 (Langmead and Salzberg 2012). After sorting and indexing the mapped reads into BAM format using samtools v1.9-45-ge415187 (Li et al. 2009), the mapped reads alignment was then used to correct the medaka-

corrected contigs using Pilon v1.23 (Walker et al. 2014) with default parameters. Since the genome was assembled using long reads and polished using both long reads and short reads, no additional scaffolding procedure was conducted explicitly, as the scaffolding information was already embedded in the long reads assembly process. The genome completeness was evaluated using BUSCO v4.0.6 (Manni et al. 2021) with insecta_odb10 library.

The genes were annotated using maker v2.31.10 (Holt and Yandell 2011). The annotation steps were briefly introduced as follows. We first identified the repetitive elements in the assembled genome using RepeatModeler v1.0.11 (http://www.repeatmasker.org/RepeatModeler/), re-annotated the "Unknown" elements into known repeat types using TEClass v2.1.3 (Abrusan et al. 2009), and quantified the genomic repeat content using RepeatMasker v4.1.1 (https://www.repeatmasker.org/RepeatMasker/). The repeat library was then specified in the maker configuration file along with the transcriptome assembled using Trinity v2.9.0 (Grabherr et al. 2011). Four runs of maker were then proceeded, in which the first run was executed with transcriptome data as a gene prediction guide, the second and third runs were conducted with trained gene models using SNAP v2.48.3 (Korf 2004), and the fourth run was performed with updated Augustus v3.3.2 gene profiles (Stanke et al. 2006).

Gene functions were predicted by mapping the predicted genes against RefSeq (Pruitt et al. 2012) (with -evalue 1e-5 cutoff) and extracting the functional annotation of the best-matched entities for the predicted genes. "Hypothetical genes" were assigned to genes that could not be mapped to any RefSeq sequences. Genome contaminations were also checked in this step, in which only contigs with the majority (> 50%) of genes mapped to Arthropoda phylum were kept for further analysis.

Data Records

Raw sequencing libraries and genome and transcriptome assemblies were deposited into NCBI SRA as part of the BioProject PRJNA699421. The nanopore reads were available under accession number SRR18531391. The BGISEQ-500RS short reads data were referred to the associated SRA and Bio-Sample numbers are SRR13626958 and SAMN17776035, respectively (Wang et al. 2021). The genome assembly with gene and transcript annotations was deposited into GenBank under the accession number JALBCR000000000. The Illumina RNA sequencing reads were available under accession number SRR20723813 (lantern) and SRR20723814 (head).

Nuclear phylogeny

Five firefly-related genomes (see Table 1) were downloaded from NCBI and other data sources. The genome of *Tribolium castaneum* (Family: Tenebrionidae) and *Ignelater luminous* (Family: Elateridae) were included as outgroups. Single copy genes were identified using orthoMCL v2.0.9 (with mcl --abc -I 1.5) (Li et al. 2003). The single copy genes were then aligned separately using Muscle v3.8.31 (Edgar 2004); poorly-aligned regions were trimmed using Gblocks v0.91b (Castresana 2000). The alignments of the single copy genes were concatenated using a custom Perl script, and the phylogenetic tree was built from the alignment using FastTree v2.1.9 (Price et al. 2010) with default parameters and visualized using MEGA X (Kumar et al. 2018).

The synonymous and non-synonymous trees were

Table 1. Species and data sources used for building the phylogenetic tree. The putative *LWS* gene candidates from each genome were identified and reannotated in the table

Species	Data source	Accession no. or website	LWS (hypothetic protein)	Reference
Lamprigera yunnana	NCBI	GCA_013368075.1	FQA39_LY08143_1402397-1401056	(Zhang et al. 2020)
Photinus pyralis	Fireflybase	http://www.fireflybase.org/	Ppyr1.3_LG5_10739432-10740780	(Fallon et al. 2018)
Aquatica lateralis	Fireflybase	http://www.fireflybase.org/	Alat1.3_scaffold_703_90391-91758	(Fallon et al. 2018)
Abscondita terminalis	NCBI	GCA_013368085.1	FQR65_LT10610_1659332-1660681	(Zhang et al. 2020)
Pyrocoelia pectoralis	Gigadb	http://gigadb.org/dataset/100376	maker-scaffold382-augustus-gene-1.421-mRNA	(Fu et al. 2017)
Outgroup:				
Ignelater luminosus	Fireflybase	http://www.fireflybase.org/	Ilumi1.2_Scaffold13785_69352-65621	(Fallon et al. 2018)
Tribolium castaneum	NCBI	GCA_000002335.3	NA	(Tribolium Genome Sequencing et al. 2008)

constructed using the codeml program of PAML v4.10.6 package (Yang 2007). The dN/dS was also estimated using PAML.

Comparative transcriptome of adult male head and lantern

To reveal tissue-specific expressed genes, we compared transcriptomes between the adult male head and lantern. The transcriptome short reads were trimmed using Trimmomatic v0.39 (Bolger et al. 2014) with the following parameters: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:12 TRAILING:12 SLIDINGWINDOW:4:15 MINLEN:36. The trimmed reads were then mapped against the assembled genome using bowtie2 v2.2.3 (Langmead and Salzberg 2012). The differential gene expression analysis was carried out by initially quantifying gene counts from the mapped reads using featureCounts v1.6.1 (Liao et al. 2014) followed by differential analysis using NOISeq v2.40.0 (Robinson and Oshlack 2010; Tarazona et al. 2015; Zhao et al. 2021) in terms of RPKM (reads per kilo-bases per million reads). Differentially-expressed genes were defined as those with at least 2 fold-of-changes. The expressed *OR*, luciferase and opsin genes in adult males were manually identified from gene annotations.

Since some genes such as the *LWS* opsin genes were incomplete in the assembled genome, an additional transcriptome quantification was conducted based on transcriptome assembly. The Trinity-assembled transcriptome was first clustered at 95% amino acid identity using CD-HIT v4.7 (Li and Godzik 2006) followed by RNA-Seq reads mapping using bowtie2 2.2.3 (Langmead and Salzberg 2012). RSEM v1.2.28 (Li and Dewey 2011) was used to quantify the expression of genes predicted by Transdecoder v5.3.0 (https://github.com/TransDecoder/TransDecoder) in terms of FPKM (fragments per kilo-bases per million reads).

Photic environment measurement

The changes of the environmental spectrum before and after firefly activities were measured using an Ocean optic USB2000+ spectrometer (Ocean Optics Inc., Dunedin, Florida, USA). The measurements were performed in an open place without the shade of tall vegetation located at the studied habitat in Nangang. During the measurement, the spectrometer probe was directly pointing towards the sky to collect the spectral range within (350–1,000 nm). Five readings were collected for each of the following time points: 1) before the fireflies started activity, 2) when the fireflies started blinking and 3) when the fireflies started flying.

LWS gene amplicon validation

Crude DNA was extracted from thoracic muscles via the ZR Tissue & Insect DNA MicroPrep™ kit (D6015). The specific primers (AcGR_fwd: 5'-ATATCAGAATGTCGGTaTTgGGT-3' and AcGR_rev: 5'-TTATGCAGTTGCTTTTTCTTCTGA-3') were designed based on available *LWS* genes of fireflies to amplify a 1348-bp segment. Polymerase chain reactions (PCRs) in 25-μL volumes were performed with a dNTP concentration of 200 μM and a primer concentration of 0.3 μM, with 50 ng of genomic DNA, one unit of TaKaRa Taq™ DNA Polymerase, and the buffer supplied by the manufacturer. The PCR was run for 30 cycles under the following conditions: denaturation at 95°C for 20 s, annealing at 55–58°C for 1 min, extension at 72°C for 2 min, and a final extension at 72°C for 5 min. The product mixture was used as a template for DNA sequencing (Mission Biotech CO., LTD, Taipei, Taiwan). The re-sequenced *LWS* gene was deposited into GenBank under accession numbers OQ225001.

RESULTS

De novo genome assembly

Table 2 summarized the genome assembly of *Abs. cerata*. Both long and short reads were sequenced from the *Abs. cerata* tissue samples, in which 2,121,726 reads with genome coverage of 10.38X were yielded from one FLO-MIN106 flowcell while 47,335,845 short reads (total size 9,467,169,000 bps) with genome coverage of 9.79X were obtained from BGISEQ-500RS. The assembly and polishing of the genome using canu v2.2 yielded 4,855 contigs (N50 = 325,269; N90 = 99,159; L50 = 907; L90 = 2,979) after the decontamination of non-firefly contigs, such as microbial genomic sequences. The assembly genome size was 967.14 Mb, with a longest contig of 2.115 Mb. 47.11% of the genome sequences consisted of repeat elements, of which DNA transposons accounted for more than half (26.83% of the genomic sequences, or 56.9% among all identified repeat).

Alternatively, the genome size could be estimated via flow cytometer (see MATERIALS AND METHODS). Table 3 shows that the genome size was about 715 Mb. The size variation between PI staining and *de novo* genome assembly might be due to insufficient staining of nuclear DNA or copy number variation of repeat elements.

Genome annotation and BUSCO analysis

127,704,450 raw reads were generated through RNA sequencing, and 126,260,884 (98.87%) clean reads were obtained after removing low quality reads. With RNA-seq based annotation to predict protein-coding genes, 55,206 protein-coding genes were predicted, and 20,862 genes could be annotated (Table S1). The genome of *Abs. cerata* is about 85% complete (BUSCO4 reported C:81.7% [S:80.5%, D:1.2%], F:3.1%, M:15.2%, n:1367 for Insecta lineage). Also, 79% of BUSCO gene completeness could be predicted by the transcriptome (C:73.9% [S:70.9%, D:3.0%], F:5.1%, M:21.0%, n:1367).

Differentially expressed genes between adult male head and lantern

We identified 12,747 out of the 20,863 annotated genes that were expressed in head and/or lantern tissues (Table S2). Among the 12,747 expressed genes, 2,076 and 2,190 of the differentially expressed genes were above two fold-of-changes in head and lantern tissues, respectively (Table S3). We wanted to know if the opsin and luciferase genes are related to nocturnal activity and mating behavior. There is one *UVS* opsin gene and one *LWS* opsin gene within the *Abs. cerata* genome. The transcriptome of the head revealed that the *UVS* opsin gene is the eighth highly expressed (FPKM = 97,932 in head; 12 in lantern) in male adults (Table 4a). However, the *LWS* opsin gene is incomplete in the genome draft due to the low coverage reads assembly and/or sequence error, with only a partial 5'-end fragment (377 bps) revealed in the 3'-end of tig00652085_segment0_pilon (84,107 kb) (Fig. 2). We then designed the specific primers to amplify the *LWS* opsin gene amplicon using genomic DNA. The complete *LWS* opsin gene amplicon (1348 bps) could be re-sequenced by Sanger sequencing. The results showed 5 exons (129–378 bps) with 4 short introns (46–55 bps). The short introns with A/T homopolymers may be the noise in the genome assembly. Since we cannot calculate those incomplete gene expressions from genome, we re-calculated the *UVS* and *LWS* opsin gene expressions via *de novo* transcriptomic assembly (Table 4b), in which the FPKM for the *LWS* and *UVS* opsin gene expression were 7,617 and 464 in head and 0 for both in lantern, respectively.

On the other hand, two ultra-high expressed genes could be found in the transcriptome of the lantern (Table 4a). The FPKM value of Luciferase 1 (*Luc1*; snap_masked-tig00000382_segment0_pilon-processed-gene-2.24) is 1,037,989 while the other FPKM value of the *Cytochrome P450 4G15*-like gene (*CYP4G15*; maker-tig00064854_segment0_pilon-snap-gene-0.23) is 805,540 in the lantern. The transcriptome of the lantern revealed highly expressed Luciferase 1 (FPKM = 2 in head; 1,037,989 in lantern) and less expressed Luciferase 2 (*Luc2*; maker-tig00008619_segment0_pilon-augustus-gene-2.43; FPKM = 18 in head; 77 in lantern) genes in male adults. The top two ultra-high expressed genes in the transcriptome of the head were arrestin-like and a small heat-shock protein (Table S3). The FPKM value of arrestin (maker-tig00014644_segment0_pilon-augustus-gene-1.39) is 929,529

Table 2. Genome assembly summary of *Abscondita cerata*

	Abscondita cerata
Assembled genome size (bp)	967,144,706
Genome coverage (fold)	10.38x
Total no of contigs	4,855
Max contig length (bp)	2,115,774
Contig N50 (bp)	325,269
Contig L50	907
Contig N90 (bp)	99,159
Contig L90	2,979
Number of genes predicted	55,206
Mean exon length (bp)	268.16
Average no of exons/gene	3.82
Insecta BUSCO gene number	1,367
% complete BUSCOs*	81.7%
% complete and single-copy BUSCOs	80.5%
% complete and duplicated BUSCOs	1.2%
% fragmented BUSCOs	3.1%
% missing BUSCOs	15.2%

* BUSCO version is 4.0.4; The lineage dataset is eukaryota_odb10 (creation date: 2019-11-20, number of species: 70, number of BUSCOs: 255).

Table 3. Genome size determination using flow cytometer of propidium iodide (PI)-stained nuclei

Species	Genome size (Mb)	Mean_PI (2N)	SD_PI (2N)
Danio rerio*	1,412	102.991	7.999
Abscondita cerata (BM1)	717 ± 64	52.272	4.647
Abscondita cerata (BM3)	712 ± 71	51.938	5.158

* zebrafish AB strain and the genome size is referred to 1,412 Mb from NCBI. Standard deviation (SD).

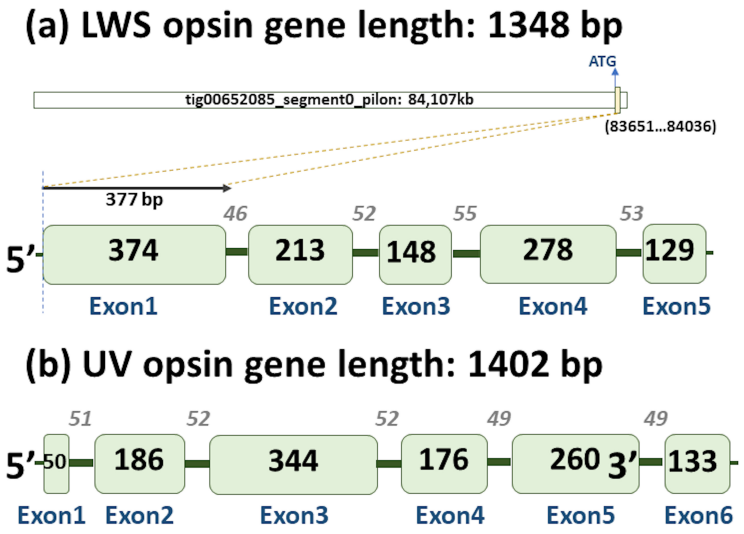

Fig. 2. Long wavelength-sensitive (*LWS*) and ultraviolet-sensitive (*UVS*) opsin gene structure. The opsin genes were amplified by degenerated primers and validated via the sanger sequencing. The exons and introns were predicted by the NCBI orffinder and supported by transcriptomic reads.

Table 4. Selected gene candidates expressed in head or lantern tissues. The expressed olfactory receptor, luciferase and opsin genes in adult males were specifically addressed in the discussion

(a) Predicted genes from genome:

Gene id	Putative gene	Head	Lantern	Fold_of	Pfam	Protein family
		(RPKM)	(RPKM)	Change		
snap_masked-tig00064992_segment0_pilon-processed-gene-1.66	**UVS opsin**	**97,932**	21	12.2	PF00001	7 transmembrane receptor
tig00652085_segment0_pilon (gDNA)*	LWS opsin	NA	NA	NA	NA	NA
snap_masked-tig00000382_segment0_pilon-processed-gene-2.24	**Luciferase 1**	2	**1,037,989**	-19.02	PF00501	AMP-binding enzyme
maker-tig00008619_segment0_pilon-augustus-gene-2.43	Luciferase 2	18	77	-2.14	PF00501	AMP-binding enzyme
maker-tig00064854_segment0_pilon-snap-gene-0.23	**CYP4G15**	7.78	**805,540**	-16.659	PF00501	Cytochrome P450
maker-tig00651154_segment0_pilon-snap-gene-0.67	**OR56d**	**18,671**	2	**13.13**	PF01395	PBP/GOBP family
snap_masked-tig00010855_segment0_pilon-processed-gene-0.13	OR19d	7,857	1,193	2.72	PF01395	PBP/GOBP family
maker-tig00651459_segment0_pilon-snap-gene-0.34	**OR56d**	7,784	4	**10.87**	PF01395	PBP/GOBP family
maker-tig00010855_segment0_pilon-augustus-gene-0.4	OR19d	4,225	353	3.58	PF01395	PBP/GOBP family
maker-tig00064231_segment0_pilon-augustus-gene-2.19	OR83a	4,073	407	3.32	PF01395	PBP/GOBP family
maker-tig00010855_segment0_pilon-snap-gene-1.0	OR19d	3,146	346	3.18	PF01395	PBP/GOBP family
maker-tig00005308_segment0_pilon-augustus-gene-2.8	**OR19d**	2,433	0	**12.29**	PF01395	PBP/GOBP family
genemark-tig00004518_segment0_pilon-processed-gene-4.15	OR99b	1,408	1,011	0.48	PF01395	PBP/GOBP family
maker-tig00005221_segment0_pilon-augustus-gene-3.75	OR59a	1,328	136	3.29		
maker-tig00065369_segment0_pilon-augustus-gene-0.64	OR83a	1,223	19	6.03	PF01395	PBP/GOBP family
maker-tig00012100_segment0_pilon-augustus-gene-0.3	**OR56d**	1,197	0	**11.26**	PF01395	PBP/GOBP family
maker-tig00652072_segment0_pilon-augustus-gene-0.6	**OR19d**	965	0	**10.95**	PF01395	PBP/GOBP family

* only 377 bp fragment could be found in genome sequence.

(b) Predicted genes from transcriptome:

Gene id	Putative gene	Head	Lantern	Pfam	Protein family
		(count, FPKM)	(count, FPKM)		
TRINITY_DN10655_c0_g1_i2	UVS opsin	(17,753, 464)	(0, 0)	PF00001	7 transmembrane receptor
TRINITY_DN15410_c3_g7_i1	LWS opsin	(3,78,270, 7,617)	(0, 0)	PF10324.8	Serpentine type 7TM GPCR chemoreceptor
TRINITY_DN16196_c1_g1_i1	Luciferase 1	(1, 0.02)	(465,842, 7,269)	PF00501.27	AMP-binding enzyme
TRINITY_DN51863_c0_g1_i1	Luciferase 2	(860, 9.40)	(1,207, 10.6)	PF00501.27	AMP-binding enzyme

while the other FPKM value of the small Heat shock protein-like gene (*Hsp20*; snap_masked-tig00003087_segment0_pilon-processed-gene-3.6) is 866,534.

In addition, there are 111 putative *OR* genes, classified into 27 *OR* gene families, within the *Abs. cerata* genome (Table S1). The transcriptome of the head revealed that 36 *OR* genes were expressed in male adults (Table 4a). The 36 *OR* genes from 15 gene families were predicted and belonged to two Pfam domains (PF01395: PBP/GOBP family and PF02949: 7tm odorant receptor). Six 7tm odorant receptors were less expressed (FPKM = 2–540 in head). For the rest of the 30 PBP/GOBP family genes, only 11 genes were specifically expressed in the head (higher FPKM > 1000 and/or FPKM = 0 in lantern), which were one *OR59a*, three *OR56d*, four *OR19d*, one *OR99b*, two *OR83a* family genes (Table 4a). "maker-tig00651154_segment0_pilon-snap-gene-0.67", an *OR56d* family gene, was the highest expressed *OR* gene, with FPKM = 18,671. The other one *OR56d* and two *OR19d* genes also showed a higher expression level than other *OR* genes.

Molecular phylogeny of single copy nuclear genes

We reconstructed the molecular phylogeny of 6 species of Lampyridae (see Table 1) based on 1,792 orthologous single copy genes (12,867,000 amino acids after alignment and Gblocks trimming) from the genomic data (Fig. 3a). Bootstrap values shown at the branch nodes indicate every clade within Lampyridae is well supported (100%). As shown in figure 3a, the family Lampyridae is a monophyletic group. The monophyly of the subfamily Luciolinae contains *Aquatica lateralis* and (*Abs. cerata* + *Abs. terminalis*) while *Lamprigera yunnana* is sister to the clade of (*Aq. lateralis* + (*Abs. cerata* + *Abs. terminalis*)). The *Photinus pyralis* and *Pyrocoelia pectoralis* form another monophyly of the subfamily Lampryriinae.

Among studied species, *Abs. cerata* is the endemic island species while others are all continental species. The phylogenetic tree suggested that the *Abs. cerata* has a slightly faster rate in proteins than that of relatively close species *Abs. terminalis*. By building both synonymous and non-synonymous trees (Fig. 3b and 3c) and estimated the dN/dS value, we estimated that the dN/dS was 0.0922, suggesting negative selection. The comparison between both the synonymous and non-synonymous tree branch lengths of *Abs. cerata* and *Abs. terminalis* also shows that the mutations in *Abs. cerata* has ~1.42 fold discrepancy than those in *Abs. terminalis*.

Photic environment

Figure 4 shows how environmental short wavelength light (< 500 nm) dramatically decreases during the twilight period, especially before and after the starting time of firefly nocturnal activity. The *Abs. cerata* start to flash or fly 5–10 minutes later when the short wavelength light disappears. In addition, the mid wavelength light (~400–600 nm) could be still detectable at this moment.

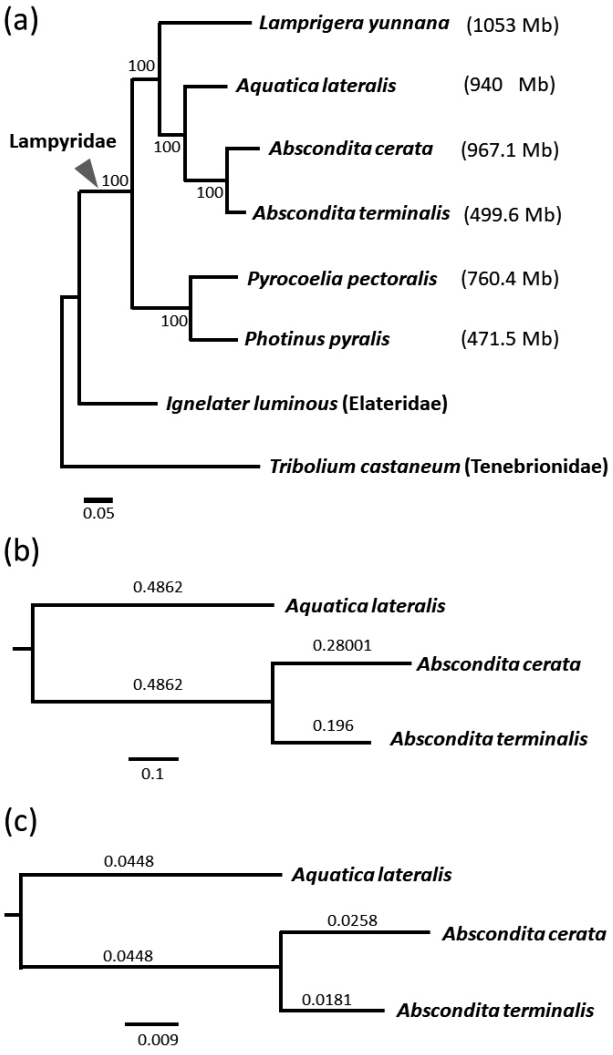

Fig. 3. Phylogenetic tree of fireflies (Lampyridae) constructed by 1,792 single copy orthologous genes. (a) protein tree, (b) synonymous tree and (c) non-synonymous tree between two *Abscondita* species. The species *Tribolium castaneum* and *Ignelater luminous* were used as the outgroups for the protein tree; the *Aquatica lateralis* species was used as the species for the synonymous and non-synonymous tree. The assembled genome size is taken from table 5.

DISCUSSION

This study demonstrated a hybrid genome assembly of an endemic firefly (*Abs. cerata* in Lampyridae) using low coverage of long-read sequences and error correcting by reasonable coverage of short-read sequences. Previous study indicated the large variations of firefly *COI* genes between Taiwan and adjacent regions (Goh et al. 2022b). Thus, the endemic *Abs. cerata* is a suitable genome reference for the molecular basis of firefly luminescence and ecology conservation.

Genome evolution of firefly

Previous studies (Liu et al. 2017; Lower et al. 2017) revealed the genome size of fireflies ranged from 410.76 to 2,154.37 Mb, which contains repeats from 10.3% to 56.6% (Lower et al. 2017). Figure 5 suggested the genome size variation may be due to different proportion of repeats (Zhang et al. 2020) and/or (tandem) gene duplications, such as acyl-CoA synthetase (*ACS*), luciferase, etc (Oba and Schultz 2022). A previous study indicated the *ACS* gene underwent a tandem duplication on a single chromosome while the neofunctionalized luciferase evolved from adjacent *ACS* and duplicated to another chromosome later (Oba and Schultz 2022). Since the *Abs. cerata* genome is still fragmented, these genes remained to separate into different contigs (data not shown). Further chromosomal-level genome will give more insight on the endemic genome evolution.

Consistent phylogeny between nuclear and mitochondrial genes

The nuclear phylogeny, constructed using 1,792 single copy genes, agreed with the other phylogenetic trees using mitochondrial genes (Chen et al. 2019; Goh et al. 2022b; Wang et al. 2021) and 436 nuclear genes (Martin et al. 2019). Figure 3a revealed the monophyly

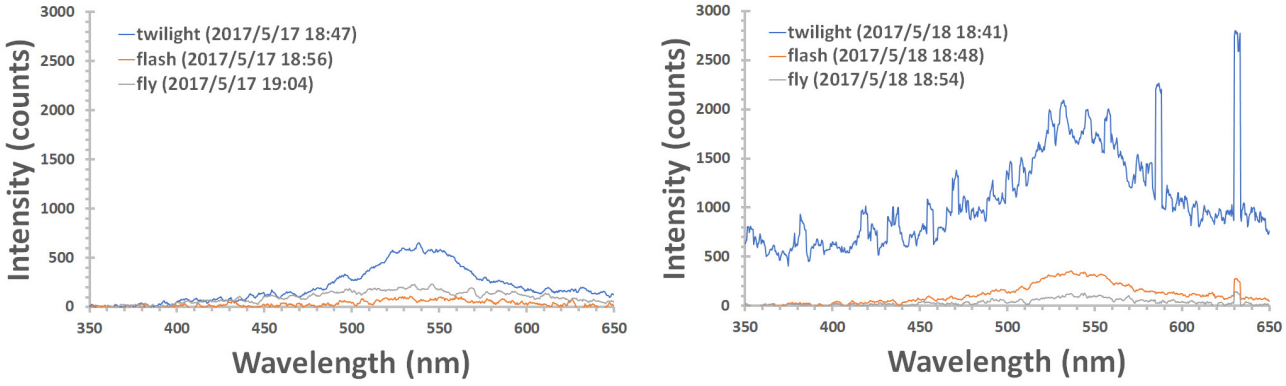

Fig. 4. Photic environment before and after the starting time of firefly nocturnal activity. After twilight begins, there is a 5–10 minute time interval between the start of firefly flash and fly conditions on two independent days (5/17 and 5/18).

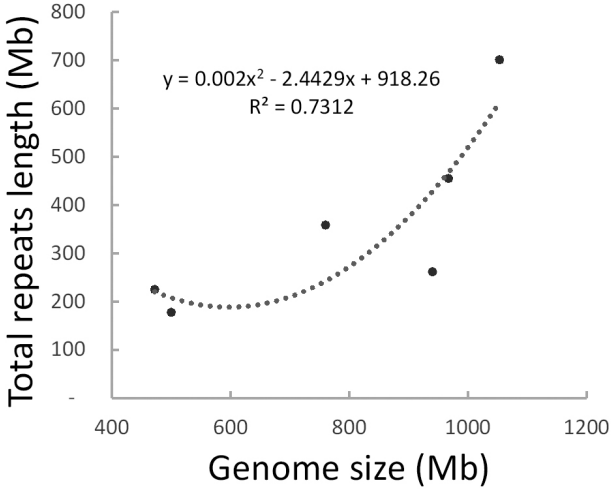

Fig. 5. Genome size and total repeats length of studied fireflies.

of the family Lampyridae, although only a few species genomes were sequenced. In addition, *Lamp. yunnana*, which was originally classified in Lampyrinae (McDermott 1966), is sister to Luciolinae with solid support in both nuclear (this study) and mitogenomic phylogenies (Wang et al. 2021). Thus, we propose that this Asian genus *Lamprigera* could be assigned to a sister group of other Luciolinae species or further considered as a unique subfamily clade in Lampyridae, which was previously suggested by morphological and *COI* barcode (Dong et al. 2021).

Island species were thought to have lower census population sizes and genetic diversity than continental species (Leroy et al. 2021). A recent study (Goh et al. 2022b) supported the low *COI* diversity in endemic *Abs. cerata* (island species). In contrast, recent GenBank sequences revealed higher *COI* diversity in the continental *Abs. chinensis* (sister species). Since the *Abs. chinensis* genome was not available, we compared the genomes of *Abs. cerata* and *Abs. terminalis*, a relatively closer species (Fig. 3b, c). Synonymous mutations and non-synonymous trees showed slightly larger discrepancy in *Abs. cerata*, which the island species may have higher mutation rate than the continental *Abs. chinensis*. In addition, their dN/dS < 1 suggests that the firefly genes may undergo negative selection and contribute to the accumulation of deleterious amino acid changes. Further comparative genomics between island and continental species could be investigated in the future.

Opsin genes revealed short wavelength sensitivity to photic environment

The spectral sensitivity of fireflies ranges from UV to red light (Martin et al. 2015). The color vision of fireflies is mainly controlled by two opsin genes, *LWS* (480–600 nm) and *UVS* (300–400 nm). The transcriptome of the head revealed that the *UVS* and *LWS* genes were both highly expressed in the eyes of male *Abs. cerata* adults (Table 4b). A previous study indicated that the Japanese firefly (*Aq. lateralis*) has a stable *UVS* gene expression in both males and females but a dynamic *LWS* gene expression in females during the diel cycle (Oba and Kainuma 2009). Both studies implied that the vision of fireflies could be more sensitive to daily shorter wavelengths via UVS opsin. Another study also mentioned that cohabitated *Aq. ficta* may change their flash signals when the light in their environment has wavelengths < 533 nm, which means the green-yellow-red light would be suitable for firefly habitats (Owens et al. 2018). All these studies revealed that the *UVS* opsin gene is what makes the vision of fireflies sensitive to environmental short wavelength light. In addition, *Abs. cerata* starts nocturnal activity slightly earlier than other cohabitated fireflies, when the environmental light intensity is lower than 6.49 lux (Goh et al. 2022ab), which might be attributed to the presence of the mid wavelength light (~400–600 nm) in figure 4. This study further reveals that the *UVS* opsin gene's major impact on firefly spectral sensitivity may lead to a coordination between the onset of flashing/nocturnal behavior and the disappearance of environmental short wavelength light (Fig. 4). Thus, different spectral sensitivities of firefly vision might be a factor that drives the starting time of nocturnal activity during evolution.

In addition, a recent study indicated a potential advantage of red sensitivity in visual discrimination of insect colors against vegetation and highlighted the potential adaptive values of long wavelength sensitivity in insects (Wang et al. 2022). Flowers and leaves were all enhanced with increasing λ_{max} from 580 nm to 640 nm (Osorio and Vorobyev 1996) while λ_{max} of the LWS photoreceptor is most commonly around 560–600 nm in insects (van der Kooi et al. 2021). Thus, the LWS opsin may enhance fireflies to sense their flash signals (λ_{max} around 552–572 nm (Goh et al. 2022b)).

The top two most expressed genes in the head were *arrestin*-like and a small heat-shock protein, *Hsp20* (Table S3). The arrestin protein family is important for regulating signal transduction at G protein-coupled receptors (Gurevich and Gurevich 2006) and desensitizing insect photoreceptor cells (Kiselev and Subramaniam 1994). The small HSP20 protein is a multifunctional protective agent (Edwards et al. 2011). These proteins are transiently upregulated in many tissue types following stressful stimuli. Both proteins may assist the environmental sensing and/or visual response in fireflies.

Olfactory receptor *OR56d* might be responsible for earthy odorant

In addition to using luminescent communication, a few fireflies may search for a male or female during courtship using pheromones (Stanger-Hall and Lloyd 2015) such as *Phausis reticulata* (De Cock et al. 2014) and *Phosphaenus hemipterus* (De Cock and Matthysen 2005). Thus, we investigated the transcriptome of the head and found 36 *OR* genes that were expressed in the *Abs. cerata* male adult (Table 4a). Among these *OR* genes, 11 head-specific expressed OR receptor genes (Table 4a) could be identified with high FPKM values. They are one *OR59a*, three *OR56d*, four *OR19d*, one *OR99b*, two *OR83a* genes, belonging to male-specific pheromone binding proteins (PBP) and/or general-odorant binding protein (GOBP) families (Vogt et al. 1991). In addition, three *OR56d* and two *OR19d* are

higher expressed genes (> 10 fold-of-change; FPKM = 965–18671) in the head of *Abs. cerata* male adults (Table 4a).

In terrestrial animals, the PBP/GOBP family of genes exist in an aqueous environment to detect odorants that are primarily hydrophobic. For example, the insect *OR56d* receptor might sense the earthy odorant geosmin, the characteristic smell of moist soil, freshly plowed earth by *Streptomyces* microbes (Jiang et al. 2007) or mold growing in rotten fruits (Gruber et al. 2018). Such a highly expressed *OR56d* gene might imply that fireflies prefer moist environments or can sense a suitable humidity for their activities. In addition, the other lower expressed pheromone receptors might imply the male and female *Abs. cerata* may not communicate with their pheromones in air during courtship. Further studies still need to confirm their pheromone olfaction with mating behavior. Nevertheless, one previous study indicated that pheromone olfaction only happened when a firefly lost its flashing ability (Stanger-Hall and Lloyd 2015). Instead of pheromone attraction, conspecific flash signals remain the dominant courtship communication method for male and female *Abs. cerata*, as previously described (Ohba and Yang 2003).

Luc1 and *CYP4G15* are two major genes highly expressed in the lantern

Two ultra-high expressed genes (*Luc1* and *CYP4G15-like*) were identified in the transcriptome of the lantern (Table 4a and Table S3). Both family genes might be important in the firefly (Fallon et al. 2018). The *CYP4G* subfamily genes function as cytochrome P450 decarbonylases, which are involved in insect cuticular hydrocarbon (CHC) synthesis (Feyereisen 2020). The CYP4G enzyme can catalyze the oxidative decarbonylation of very-long chain aldehydes to cuticular hydrocarbons (CHCs) or directly reduce alcohols to aldehydes and form the waxy layer on the cuticle (Holze et al. 2021). The highly expressed *CYP4G15-like* gene in the lantern might imply that CHCs form a waxy layer on the cuticle of the insect or act as signaling molecules in mate recognition and chemical communication (Chung and Carroll 2015; Holze et al. 2021). There still might be some unknown reasons why the *CYP4G15-like* gene is highly expressed in the lantern. The *Luc1*-type luciferase gene is the major expressed gene responsible for the luminescent emission in the male adult, though the *Luc2*-type luciferase gene is also expressed, but with very low levels in the *Abs. cerata* lantern (Table 4a). Thus, the *Luc1*-type luciferase may contribute most to the light production from the lanterns of the *Abs. cerata* adult. *Luc1*-type luciferase is often responsible for the "yellowish" luminescence of adult firefly lanterns (Bessho-Uehara et al. 2017), and the λ_{max} of *Abs. cerata* male lanterns (563.6 ± 0.3 nm) is also a "yellowish" luminescence (Goh et al. 2022b). In addition, there might be some spectral shift due to environmental changes. For example, a previous study revealed various λ_{max} of *in vitro Pyrocoelia atripennis Luc1*-type luciferase ranged from 547–570 nm under different pH conditions (Bessho-Uehara et al. 2017). Further biochemistry of *in vitro Luc1*-type luciferase could reveal the differences.

Long-reads sequencing technology in insects

This study demonstrates that shallow sequencing of Nanopore long reads (about 10X) followed by short reads polishing may achieve similar firefly genome assembly results as other, deeper long reads sequencing firefly genome projects (with sequencing depths range from 40X to 75X), as shown in table 5. This indicates that the economical hybrid assembly strategy shown in this study could save both time and money and may

Table 5. Genome assembly comparison of *Abscondita cerata* (this study) and related fireflies

	Abscondita cerata	*Abscondita terminalis*	*Aquatica lateralis*	*Lamprigera yunnana*	*Photinus pyralis*	*Pyrocoelia pectoralis*
Assembled genome size (bp)	967,144,706	499,652,588	940 Mbps	1,052,929,944	471,511,253	760,416,098
Sequencing technology	Nanopore R9; BGISEQ-500RS	PacBio RSII	Illumina; PacBio sequal	PacBio RSII	PacBio RSII; Illumina HiSeq	PacBio RSII; Illumina HiSeq
Genome coverage (fold)	10.38x	67.39x	64x	74.05x	40.0x	~75x
Assembly method	Canu v2.2	wtdbg v. 2	MaSuRCA v. MAY-2016	wtdbg v. 2	MaSuRCA v. MAY-2016	Falcon v0.4
Total no of contigs	4,855	3,187	5,365	2,684	2,160	474
Max contig length (bp)	2,115,774	2,102,812	3,615,483	7,468,035	70,905,644	13,688,299
Contig N50 length (bp)	325,269	1,220,005	694,267	3,509,824	47,017,841	3,035,809
Contig L50	907	105	381	79	5	79

potentially accelerate insect genome projects. The gene analysis results also revealed that the genes predicted from this genome project may provide insights into firefly behavior, luminescence evolution and ecology conservation.

CONCLUSIONS

This study demonstrates that low coverage Nanopore long reads polished with BGI PE100 short reads may cover ~85% of genes in a drafted genome. The *Abs. cerata* genome and transcriptome reveals (1) their vision is mainly *UVS*, which they use to sense environmental twilight and decide when to begin nocturnal activity, (2) the major expressed olfactory receptors (*OR56d*) may be correlated to the suitable humidity of their activity, and (3) *Luc1*-type luciferase is responsible for *Abs. cerata* luminescent spectrum. Further analysis will reveal more insight on their physiological behavior, cohabitated species competition and adaptation to the specific environments.

Acknowledgments: This study is part of the Taiwan Biogenome Project, which is funded by Academia Sinica, Taiwan. We thank Mr. Shih-Chieh Huang for providing firefly photo. Thanks also to Justin Pelofsky of Third Draft Editing for his English language editing.

Authors' contributions: Conceptualization, T.-Y.W., K.-S.G., and F.-Y. W.; data curation, T.-Y.W., K.-S.G., and Y.-W.W.; methodology, T.-Y.W., K.-S.G., Y.-W.W., and L.-L.W.; project administration, T.-Y.W. and Y.-W. W.; writing–original draft, T.-Y.W., K.-S.G., L.-J.W., F.-Y. W. and Y.-W.W.; writing–review and editing, T.-Y.W., K.-S.G., L.-J.W., L.-L.W, F.-Y. W. and Y.-W.W.

Competing interests: All authors declare that they have no conflict of interest.

Availability of data and materials: Raw sequencing libraries and genome and transcriptome assemblies are deposited at NCBI SRA as part of the BioProject PRJNA699421. The nanopore DNA sequencing reads are available under accession number SRR18531391. The genome assembly with gene and transcript annotations has been deposited at GenBank under the accession number JALBCR000000000. The Illumina RNA sequencing reads are available under accession number SRR20723813 and SRR20723814. The re-sequenced *LWS* gene was deposited into GenBank under accession numbers OQ225001.

Consent for publication: Not applicable.

Ethics approval consent to participate: Not applicable.

REFERENCES

Abrusan G, Grundmann N, DeMester L, Makalowski W. 2009. TEclass - a tool for automated classification of unknown eukaryotic transposable elements. Bioinformatics **25:**1329–1330. doi:10.1093/bioinformatics/btp084.

Ballantyne L, Fu XH, Lambkin C, Jeng ML, Faust L, Wijekoon WMCD et al. 2013. Studies on South-east Asian fireflies: *Abscondita*, a new genus with details of life history, flashing patterns and behaviour of *Abs. chinensis* (L.) and *Abs. terminalis* (Olivier) (Coleoptera: Lampyridae: Luciolinae). Zootaxa **3721:**1–48. doi:10.11646/zootaxa.3721.1.1.

Ballantyne LA, Lambkin CL, Ho JZ, Jusoh WFA, Nada B, Nak-Eiam S et al. 2019. The Luciolinae of S. E. Asia and the Australopacific region: a revisionary checklist (Coleoptera: Lampyridae) including description of three new genera and 13 new species. Zootaxa **4687:**1–174. doi:10.11646/zootaxa.4687.1.1.

Bessho-Uehara M, Konishi K, Oba Y. 2017. Biochemical characteristics and gene expression profiles of two paralogous luciferases from the Japanese firefly *Pyrocoelia atripennis* (Coleoptera, Lampyridae, Lampyrinae): insight into the evolution of firefly luciferase genes. Photochem Photobiol Sci **16:**1301–1310. doi:10.1039/c7pp00110j.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30:**2114–2120. doi:10.1093/bioinformatics/btu170.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol **17:**540–552. doi:10.1093/oxfordjournals.molbev.a026334.

Chen X, Dong Z, Liu G, He J, Zhao R, Wang W et al. 2019. Phylogenetic analysis provides insights into the evolution of Asian fireflies and adult bioluminescence. Mol Phylogenet Evol **140:**106600. doi:10.1016/j.ympev.2019.106600.

Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S et al. 2018. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. Gigascience **7:**1–6. doi:10.1093/gigascience/gix120.

Chung H, Carroll SB. 2015. Wax, sex and the origin of species: Dual roles of insect cuticular hydrocarbons in adaptation and mating. Bioessays **37:**822–830. doi:10.1002/bies.201500014.

De Cock R, Faust L, Lewis S. 2014. Courtship and mating in *Phausis reticulata* (Coleoptera: Lampyridae): male flight behaviors, female glow displays, and male attraction to light traps. Fla Entomol **97:**1290–1307. doi:10.1653/024.097.0404.

De Cock R, Matthysen E. 2005. Sexual communication by pheromones in a firefly, *Phosphaenus hemipterus* (Coleoptera: Lampyridae). Anim Behav **70:**807–818. doi:10.1016/j.anbehav.2005.01.011.

Dong ZW, Yiu V, Liu GC, He JW, Zhao RP, Peng YQ et al. 2021. Three new species of *Lamprigera* Motschulsky (Coleoptera, Lampyridae) from China, with notes on known species. Zootaxa **4950:**441–468. doi:10.11646/zootaxa.4950.3.2.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32:**1792–1797. doi:10.1093/nar/gkh340.

Edwards HV, Cameron RT, Baillie GS. 2011. The emerging role of HSP20 as a multifunctional protective agent. Cell Signal **23:**1447–1454. doi:10.1016/j.cellsig.2011.05.009.

Fallon TR, Lower SE, Chang CH, Bessho-Uehara M, Martin GJ, Bewick AJ et al. 2018. Firefly genomes illuminate parallel

origins of bioluminescence in beetles. Elife **7:**e36495. doi:10.7554/eLife.36495.

Feyereisen R. 2020. Origin and evolution of the CYP4G subfamily in insects, cytochrome P450 enzymes involved in cuticular hydrocarbon synthesis. Mol Phylogenet Evol **143:**106695. doi:10.1016/j.ympev.2019.106695.

Fu X, Li J, Tian Y, Quan W, Zhang S, Liu Q et al. 2017. Long-read sequence assembly of the firefly *Pyrocoelia pectoralis* genome. Gigascience **6:**1–7. doi:10.1093/gigascience/gix112.

Goh KS, Lee CM, Wang TY. 2022a. Species-specific flash patterns track the nocturnal behavior of sympatric Taiwanese fireflies. Biology-Basel **11:**58. doi:10.3390/biology11010058.

Goh KS, Li CW. 2011. A photocytes - associated fatty acid-binding protein from the light organ of adult Taiwanese firefly, *Luciola cerata*. PLoS ONE **6:**e29576. doi:10.1371/journal.pone.0029576.

Goh KS, Wang LJ, Ni JH, Wang TY. 2022b. Luminescent characteristics and mitochondrial *COI* barcodes of nine cohabitated Taiwanese fireflies. PeerJ **10:**e14195. doi:10.7717/peerj.14195.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol **29:**644–652. doi:10.1038/nbt.1883.

Gruber L, Rybak J, Hansson BS, Cantera R. 2018. Synaptic spinules in the olfactory circuit of *Drosophila melanogaster*. Front Cell Neurosci **12:**86. doi:10.3389/fncel.2018.00086.

Gurevich EV, Gurevich VV. 2006. Arrestins:ubiquitous regulators of cellular signaling pathways. Genome Biol **7:**236. doi:10.1186/gb-2006-7-9-236.

Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics **12:**491. doi:10.1186/1471-2105-12-491.

Holze H, Schrader L, Buellesbach J. 2021. Advances in deciphering the genetic basis of insect cuticular hydrocarbon biosynthesis and variation. Heredity **126:**219–234. doi:10.1038/s41437-020-00380-y.

Jeng ML, Lai J, Yang PS. 1999. A synopsis of the firefly fauna at six national parks in Taiwan (Coleoptera: Lampyridae). Formosan Entomologist **19:**65–91.

Jeng ML, Yang PS, Sato M, Lai J, Chang JC. 1998. The genus *Curtos* (Coleoptera, Lampyridae, Luciolinae) of Taiwan and Japan. The Japanese Journal of Systematic Entomology **4:**331–347.

Jiang J, He X, Cane DE. 2007. Biosynthesis of the earthy odorant geosmin by a bifunctional *Streptomyces coelicolor* enzyme. Nat Chem Biol **3:**711–715. doi:10.1038/nchembio.2007.29.

Kiselev A, Subramaniam S. 1994. Activation and regeneration of rhodopsin in the insect visual cycle. Science **266:**1369–1373. doi:10.1126/science.7973725.

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res **27:**722–736. doi:10.1101/gr.215087.116.

Korf I. 2004. Gene finding in novel genomes. BMC Bioinformatics **5:**59. doi:10.1186/1471-2105-5-59.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol **35:**1547–1549. doi:10.1093/molbev/msy096.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods **9:**357–359. doi:10.1038/nmeth.1923.

Leroy T, Rousselle M, Tilak MK, Caizergues AE, Scornavacca C, Recuerda M et al. 2021. Island songbirds as windows into evolution in small populations. Curr Biol **31:**1303–1310.e4. doi:10.1016/j.cub.2020.12.040.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics **12:**323. doi:10.1186/1471-2105-12-323.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics **34:**3094–3100. doi:10.1093/bioinformatics/bty191.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N et al. 2009. The sequence alignment/map format and SAMtools. Bioinformatics **25:**2078–2079. doi:10.1093/bioinformatics/btp352.

Li L, Stoeckert Jr CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res **13:**2178–2189. doi:10.1101/gr.1224503.

Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics **22:**1658–1659. doi:10.1093/bioinformatics/btl158.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics **30:**923–930. doi:10.1093/bioinformatics/btt656.

Liu GC, Dong ZW, He JW, Zhao RP, Wang W, Li XY. 2017. Genome size of 14 species of fireflies (Insecta, Coleoptera, Lampyridae). Zool Res **38:**449–458. doi:10.24272/j.issn.2095-8137.2017.078.

Lower SS, Johnston JS, Stanger-Hall KF, Hjelmen CE, Hanrahan SJ, Korunes K et al. 2017. Genome size in North American fireflies: substantial variation likely driven by neutral processes. Genome Biol Evol **9:**1499–1512. doi:10.1093/gbe/evx097.

Manni M, Berkeley MR, Seppey M, Simao FA, Zdobnov EM. 2021. 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. doi:10.1093/molbev/msab199.

Martin GJ, Branham MA, Whiting MF, Bybee SM. 2017. Total evidence phylogeny and the evolution of adult bioluminescence in fireflies (Coleoptera: Lampyridae). Mol Phylogenet Evol **107:**564–575. doi:10.1016/j.ympev.2016.12.017.

Martin GJ, Lord NP, Branham MA, Bybee SM. 2015. Review of the firefly visual system (Coleoptera: Lampyridae) and evolution of the opsin genes underlying color vision. Org Divers Evol **15:**513–526. doi:10.1007/s13127-015-0212-z.

Martin GJ, Stanger-Hall KF, Branham MA, Da Silveira LFL, Lower SE, Hall DW et al. 2019. Higher-level phylogeny and reclassification of Lampyridae (Coleoptera: Elateroidea). Insect Syst Diver **3(6):**11. doi:10.1093/isd/ixz024.

McDermott FA. 1966. Pars 9 (Editio secunda). Lampyridae. *In*: W. O. Steel (ed.), Coleopterorum Catalogus Supplementa. W. Junk, The Hague, 149 pp.

Oba Y, Furuhashi M, Bessho M, Sagawa S, Ikeya H, Inouye S. 2013. Bioluminescence of a firefly pupa: involvement of a luciferase isotype in the dim glow of pupae and eggs in the Japanese firefly, *Luciola lateralis*. Photochem Photobiol Sci **12:**854–863. doi:10.1039/c3pp25363e.

Oba Y, Kainuma T. 2009. Diel changes in the expression of long wavelength-sensitive and ultraviolet-sensitive opsin genes in the Japanese firefly, *Luciola cruciata*. Gene **436:**66–70. doi:10.1016/j.gene.2009.02.001.

Oba Y, Schultz DT. 2022. Firefly genomes illuminate the evolution of beetle bioluminescent systems. Curr Opin Insect Sci **50:**100879. doi:10.1016/j.cois.2022.100879.

Ohba N, Yang PS. 2003. Flash patterns and communication system of the Taiwan firefly, *Luciola cerata* Olivier. Science Report of the Yokosuka City Museum **50:**1–12.

Orgogozo V, Rockman MV. 2011. Molecular methods for evolutionary genetics. Humana Press, New York, USA.

Osorio D, Vorobyev M. 1996. Colour vision as an adaptation to frugivory in primates. P Roy Soc B-Biol Sci **263:**593–599. doi:10.1098/rspb.1996.0089.

Owens ACS, Lewis SM. 2018. The impact of artificial light at night on nocturnal insects: A review and synthesis. Ecol Evol **8:**11337–11358. doi:10.1002/ece3.4557.

Owens ACS, Meyer-Rochow VB, Yang EC. 2018. Short- and mid-wavelength artificial light influences the flash signals of *Aquatica ficta* fireflies (Coleoptera: Lampyridae). PLoS ONE **13:**e0191576. doi:10.1371/journal.pone.0191576.

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 - approximately maximum-likelihood trees for large alignments. PLoS ONE **5:**e9490. doi:10.1371/journal.pone.0009490.

Pruitt KD, Tatusova T, Brown GR, Maglott DR. 2012. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Res **40:**D130–135. doi:10.1093/nar/gkr1079.

Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol **11:**R25. doi:10.1186/gb-2010-11-3-r25.

Sharkey CR, Fujimoto MS, Lord NP, Shin S, McKenna DD, Suvorov A et al. 2017. Overcoming the loss of blue sensitivity through opsin duplication in the largest animal group, beetles. Sci Rep-Uk **7:**8. doi:10.1038/s41598-017-00061-7.

Stanger-Hall KF, Lloyd JE. 2015. Flash signal evolution in *Photinus* fireflies: character displacement and signal exploitation in a visual communication system. Evolution **69:**666–682. doi:10.1111/evo.12606.

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: *ab initio* prediction of alternative transcripts. Nucleic Acids Res **34:**W435–439. doi:10.1093/nar/gkl200.

Tarazona S, Furio-Tari P, Turra D, Pietro AD, Nueda MJ, Ferrer A et al. 2015. Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. Nucleic Acids Res **43:**e140. doi:10.1093/nar/gkv711.

Tribolium Genome Sequencing C, Richards S, Gibbs RA, Weinstock GM, Brown SJ, Denell R et al. 2008. The genome of the model beetle and pest *Tribolium castaneum*. Nature **452:**949–955. doi:10.1038/nature06784.

van der Kooi CJ, Stavenga DG, Arikawa K, Belusic G, Kelber A. 2021. Evolution of insect color vision: from spectral sensitivity to visual ecology. Annu Rev Entomol **66:**435–461. doi:10.1146/annurev-ento-061720-071644.

Vaser R, Sovic I, Nagarajan N, Sikic M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res **27:**737–746. doi:10.1101/gr.214270.116.

Vogt RG, Prestwich GD, Lerner MR. 1991. Odorant-binding-protein subfamilies associate with distinct classes of olfactory receptor neurons in insects. J Neurobiol **22:**74–84. doi:10.1002/neu.480220108.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE **9:**e112963. doi:10.1371/journal.pone.0112963.

Wang LJ, Wu YW, Wang TY. 2021. Characterization of the complete mitochondrial genome of *Abscondita cerata* (Olivier, 1911) (Coleoptera: Lampyridae) and its phylogenetic implications. Mitochondrial DNA B Resour **6:**2528–2530. doi:10.1080/23802359.2021.1959456.

Wang LY, Stuart-Fox D, Walker G, Roberts NW, Franklin AM. 2022. Insect visual sensitivity to long wavelengths enhances colour contrast of insects against vegetation. Sci Rep **12:**982. doi:10.1038/s41598-021-04702-w.

Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. Microb Genom **3:**e000132. doi:10.1099/mgen.0.000132.

Wu CH, Jeng ML, South A, Ho JZ, Yang PS. 2010. Evidence for two male morphs of *Luciola cerata* Olivier (Coleoptera: Lampyridae) exhibiting distinct mating behavior, with implications for sexual selection. Coleopts Bull **64:**235–242. doi:10.1649/0010-065X-64.3.235.9.

Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol Biol Evol **24:**1586–1591. doi:10.1093/molbev/msm088.

Zhang R, He J, Dong Z, Liu G, Yin Y, Zhang X et al. 2020. Genomic and experimental data provide new insights into luciferin biosynthesis and bioluminescence evolution in fireflies. Sci Rep **10:**15882. doi:10.1038/s41598-020-72900-z.

Zhao Y, Li MC, Konate MM, Chen L, Das B, Karlovich C et al. 2021. TPM, FPKM, or normalized counts? A comparative study of quantification measures for the analysis of RNA-seq data from the NCI patient-derived models repository. J Transl Med **19:**269. doi:10.1186/s12967-021-02936-w.

Supplementary materials

Table S1. List of the 55,206 predicted protein-coding genes. 20,862 genes could be functionally annotated, with the exceptions of the “hypothetic protein”. (download)

Table S2. 12,747 genes were expressed in head and/or lantern tissues. The RPKM represented the gene expression level. (download)

Table S3. Differentially expressed genes with at least 2 fold-of-change between head and lantern tissues. (download)