The First Chromosome-level Genome Assembly of *Cheumatopsyche charites* Malicky and Chantaramongkol, 1997 (Trichoptera: Hydropsychidae) Reveals How It Responds to Pollution

Xinyu Ge 🟢, Jianfeng Jin, Lang Peng, Haoming Zang, Beixin Wang*, and Changhai Sun*

Department of Entomology, College of Plant Protection, Nanjing Agricultural University, 210095, P. R. China

*Corresponding authors: E-mails: wangbeixin@njau.edu.cn (B.W.); chsun@njau.edu.cn (C.S.).

Abstract

Trichoptera is a highly adapted group of freshwater insects. They are generally more sensitive to dissolved oxygen and water quality than most freshwater organisms, and this sensitivity allows them to be used as reliable biological indicators of water quality. At present, there exists no chromosome-level genome of a hydropsychid species. *Cheumatopsyche charites* Malicky & Chantaramongkol, 1997 can successfully survive and thrive in polluted streams where other caddisflies are infrequent, suggesting that they are tolerant to latent contamination. Here we report a high-quality chromosome-level genome assembly of *C. charites* generated combining PacBio long reads and Hi-C reads. We obtained a genome assembly of 223.23 Mb, containing 68 scaffolds with an N50 length of 13.97 Mb, and 155 contigs (99.67%) anchored into 16 pseudochromosomes. We identified 36.12 Mb (16.18%) of the genome as being composed of repetitive elements, identified 369 noncoding RNAs, and predicted 8,772 protein-coding genes (96.80% BUSCO completeness). Gene family evolution analyses identified 7,148 gene families, of which 41 experienced rapid evolution. The expanded gene families were shown to be involved in detoxification metabolism, digestive absorption, and resistance to viruses or bacteria. This high-quality genome provides a valuable genomic basis for the study of trichopteran evolution.

Key words: caddisflies, comparative genomics, gene family evolution, adaptation.

Significance

Caddisflies are a significant group of aquatic insects. The species diversity within this group is greater than the total of all whole aquatic insect orders. However, to date, 23 trichopteran genomes have been published, but no chromosome-level hydropsychid genome has yet been produced. The lack of such a genome restricts our understanding of trichopteran biological behavior and evolution. Here, we constructed a chromosome-level genome assembly of *Cheumatopsyche charites* and performed sex chromosome identification and gene family evolution analysis. Our results will improve our understanding of trichopteran evolution and will provide a valuable reference genome for future comparative genomic studies.

Introduction

Trichoptera (caddisfly) is an order of the Insecta containing more than 16,000 extant species, which is the seventh-highest number of any such order (Holzenthal et al. 2011; Jonason et al. 2017; Clair et al. 2018). Moreover, caddisflies among the most successful insect lineages and are often particularly important and ubiquitous members of aquatic communities in both running (lotic) and standing (lentic)

water bodies (Wiggins 1996; Holzenthal et al. 2007; Morse et al. 2019).

The cosmopolitan family Hydropsychidae (Curtis 1835) is the third largest family of Trichoptera (Geraci et al. 2005; Ge et al. 2020). The hydropsychine genus *Cheumatopsyche* (Wallengren 1891) is represented on every continent except South America, and includes nearly 390 species (Morse 2022). These species are known for their wide distribution, being present in all biogeographical regions except the Neotropics and Antarctica (Oláh et al. 2008). *Cheumatopsyche* larvae tend to be dominant in warmer streams, as they have a considerably high stain resistance; they are also often present in polluted streams where other caddisflies are usually absent (Wiggins 1996). Calculated by biomass *Cheumatopsyche* is the dominant genus in the order Trichoptera, and larval population densities can reach up to 250,000 individuals per square meter (Gibbs 1973; Botosaneanu et al. 1990). The food of *Cheumatopsyche* larvae mainly consists of algae with a small detrital component (Coffman et al. 1971). Their eating habits allow the body to remove large amounts of suspended solids in all kinds of running freshwater, and may also assist in maintaining their purification capacity (Morse et al. 2019). Despite the importance of *Cheumatopsyche*, to date no high-quality chromosomal assembly is available. At present, 23 trichopteran genomes have been published, of which two are for hydropsychine species (Heckenhauer et al. 2019, 2022).

In this study, we report the first chromosome-level Trichoptera genome assembly (*Cheumatopsyche charites*, Malicky and Chantaramongkol 1997), and report the identification of the Z chromosome. We also annotated all predicted repetitive elements, noncoding RNAs (ncRNAs), and protein-coding genes (PCGs). Using this data, we reconstructed the phylogenetic relationships between *C. charites* and seven other insect orders and analyzed the evolution of key gene families. The construction of a chromosome-level genome for *C. charites* is an important preliminary step for future genomic and applied research on caddisflies, and may also enable further comparative studies of their evolution and biological behavior.

Results and Discussion

Genome Survey and Assembly

In total, we generated 148.57 Gb clean reads for both the genome and transcriptome assemblies. These included 37.42 Gb (168×) PacBio long reads, 60.55 Gb (male 26.01 Gb 117×, female 34.54 Gb 154×) Illumina short reads, 43.97 Gb (197×) Hi-C reads, and 6.63 Gb transcriptome reads. After quality control, we retained 26.01 Gb Illumina reads of the male adult specimen and 32.37 Gb

Illumina reads of the female adult specimen for further genomic mapping, polishing, and subsequent identification of the sex chromosome. Our genome survey suggested that the *C. charites* had a genome size of 218.9–220.2 Mb. Heterozygotic analyses revealed that our *C. charites* specimens had relatively low heterozygosity (0.83–0.88%).

Flye was used to create a preliminary assembly of long reads, resulting in a 225.08 Mb assembly with an N50 length of 2.85 Mb. After polishing, redundancy checks, and Hi-C scaffolding and contaminant removal, we then obtained an assembled *C. charites* genome of 223.233 Mb. The length of the 16 identifiable pseudochromosomes was relatively even, ranging from 10.35 to 16.01 Mb (fig. 1a). The autosomal average sequencing depth (approximately 140×) was significantly higher than the sequencing depth for the sex (Z) chromosome (approximately 76.39×). Our assembly achieved a high completeness score of 99.40% (table 1), as we obtained high mapping rates of both Illumina (97.23%) and PacBio (97.61%) reads. The size of two comparable hydropsychid genomes (*Hydropsyche tenuis*, 229.66 Mb and *Parapsyche elsis*, 282.18 Mb) was both slightly greater than *C. charites*. Moreover, compared with extant public chromosome-level trichopteran genomes, the BUSCO completeness score was significantly higher for *C. charites*, suggesting that the obtained assembly was high quality (table 1).

Genome Annotation

RepeatMasker identified a total of 36.12 Mb (16.18%) of the *C. charites* genome as repetitive elements. Of these, unclassified elements represented the largest proportion (10.98%). This phenomenon may be due to insufficient studies of the repetitive elements in Trichoptera species. Other less common repetitive elements included DNA transposons (3.24%), long-terminal repeat elements (LTR, 0.18%), long-interspersed elements (LINE, 0.17%), and simple repeats (1.15%; fig. 1b; Supplementary table S1, Supplementary Material online). The proportion of repetitive elements was significantly less than those present in other trichopteran genomes (Heckenhauer et al. 2022). In addition, we also identified 369 ncRNA sequences, including 29 ribosomal RNAs, 52 microRNAs, 31 small nuclear RNAs, one long ncRNA, two ribozymes, 32 other ncRNAs, and 222 tRNAs (Supplementary table S2, Supplementary Material online).

Next, we used the MAKER3 pipeline to identify 8,772 PCGs in the *C. charites* genome. Identified PCGs were found to have a mean length of 5,974.8 bp, while the average exon number per gene was 12.5. The BUSCO completeness of the PCGs was 96.8% ($n = 1,367$). Diamond searches aligned 8,699 (99.17%) of identifiable genes to the UniProtKB database. We then functionally annotated 7,963 Gene Ontology (GO) items, 6,787 KEGG pathway
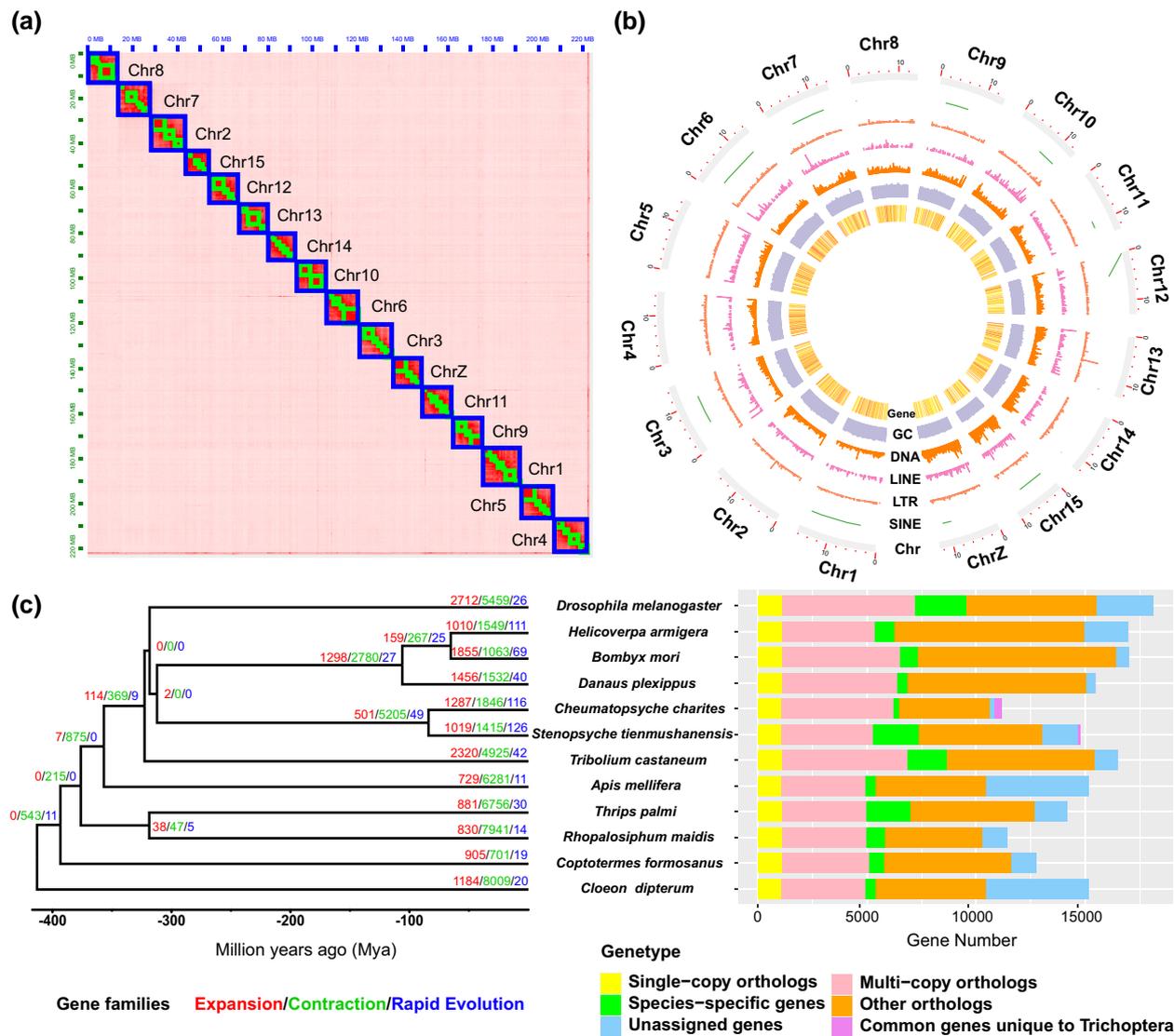
FIG. 1—Landscape, characteristics and phylogeny of genome of *Cheumatopsyche charites*. (*a*) The genome-wide Hi-C contact maps of 16 chromosomes in *C. charites*. (*b*) Characteristics map showing each chromosome characters of *C. charites* protein-coding genes density (Gene), guanine–cytosine content (GC), DNA transposons (DNA), long-interspersed elements (LINE), long-terminal repeat elements (LTR), short-interspersed elements (SINE), each chromosome length (Chr). (*c*) Phylogeny and gene family analysis of *C. charites*.

items (combining InterProScan and eggNOG results), 5,943 MetaCyc items, 7,568 Reactome items, and 8,410 COG Functional Categories.

To reveal the evolutionary relationships among members of various gene families, we used OrthoFinder to identify and characterize gene families among 12 species in eight orders. In total, 160,141 genes were assigned into 15,534 orthogroups (gene families), of which 3,968 orthogroups were present in all species. We also found that 3,521 orthogroups and 16,061 genes were species specific, and 1,108 genes were single copy (fig. 1c). We used orthologous single-copy genes to construct a phylogenetic tree. About 172 single-copy genes were filtered

by IQ-TREE using the "symtest" function. The resulting 936 loci were used to construct a phylogenetic tree. The resulting phylogeny was highly consistent with those reported by Misof et al. (2014) and Wipfler et al. (2019), and confirmed the monophyly of Trichoptera. Our results also agreed that Trichoptera originated during the late Carboniferous period (312–314 Ma) as a sister clade of Lepidoptera, as well as that the divergence point between the Stenopsychidae and the Hydropsychidae occurred in the early Jurassic period (180 Ma). These observations agreed with earlier results (Thomas et al. 2020).

Based on this phylogeny, we conducted gene family expansion and contraction analyses. First, 10,992 genes from

Table 1

Assembly Statistic Comparison of Hydropsychidae Genomes

	Cheumatopsyche charites[a]	Limnephilus marmoratus	Glyphotaelius pellucidus
Accession no.	JACZEN000000000	GCA_009617725.1	JAGVSN000000000
Assembly size (Mb)	223.23	1629.97	1037.14
Number of scaffolds/ contigs	68/207	68/395	57/285
Longest scaffold/contig (Mb)	16.52/6.87	77.26/28.54	51.24/43.27
N50 scaffold/contig length (Mb)	13.97/2.85	56.17/8.02	36.81/8.16
GC (%)	32.97	35.73	35.10
Gaps (%)	0.01	0.007	0.004
BUSCO completeness (%)[b]	C: 99.4% (S: 99.0%, D: 0.4%), F: 0.3%, M: 0.3%	C: 97.8% (S: 96.9%, D: 0.9%), F: 0.3%, M: 1.9%	C: 98.1% (S: 97.2%, D: 0.59%), F: 0.1%, M: 1.8%

C, complete; S, single; D, duplicated; F, fragmented; M, missing.

[a]Assemblies produced in this study.

[b]$N_{Insecta} = 1,367$.

the whole *C. charites* genome assembly were clustered into 7,148 gene families; of these, 1,287 had expanded and 1,846 had contracted. Of these gene families, 41 had experienced significant expansion ($P < 0.05$), and these included cytochrome P450, trypsin, and cuticle protein cuticle protein A3A. Analyses from previous reports aligned with our data and showed that these gene families were involved in processes such as detoxification metabolism, digestive absorption, and resistance to infection by foreign viruses or bacteria. We therefore speculated that large expansions of the immune-, digestion-, and detoxification metabolism–related gene families were crucial for the adaptations of caddisflies, especially as larvae, to withstand polluted and harsh environments. With deeper sampling and transcriptomic analyses of different instars in future, we will better understand which specific genes may be involved in the adaptation of hydropsychidan species to survival in relatively contaminated environments. Simultaneously, new high-quality genomes of Hydropsychidae species may provide valuable information that will enable us to better interpret the phylogeny of the Trichoptera. In addition, the identification of the trichopteran Z chromosome may serve as an important reference for further analysis of the particular mechanisms involved in Z0 sex determination, both in the Trichoptera as well as in the Insecta at large.

Materials and Methods

Sample Collection and Sequencing

Adult specimens of *C. charites* were collected using pan traps with 15 W ultraviolet light tubes on May 19, 2020, Longchuan river, Yuanmou County (25.961°N, 101.874°E; Alt: 880.1 m), Yunnan province, China. Samples were flash frozen and stored at −80 °C before extraction. All samples were washed with ddH$_2$O and intestines were removed (to minimize gut microbial contamination) before samples were sent for sequencing. Specimen identification was performed by X-Y Ge and C-H Sun, as well as by integrated morphology and COI barcode. One male and one female specimen were used for Illumina whole-genome sequencing. One male individual was used for Illumina transcriptomic sequencing and Hi-C sequencing. Ten male specimens were used for PacBio long-read sequencing. Genomic DNA and RNA extraction, short-read library construction, and DNA digestion procedures followed those reported by Wang et al. (2022). A long-read library with a 15-kb-insert size was generated on the PacBio Sequel II platform. All library production and sequencing were generated at Berry Genomics.

Genome Survey and Assembly

BBTools v38.29 (Bushnell 2014) was used for second-generation quality control of the raw Illumina sequence data. Quality control included several processes, such as removal of duplicates (with "clumpify.sh"), adapter trimming (using "bbduk.sh"), quality trimming (>Q20), polymer trimming (>10 bp for poly-A/G/C tails), length filtering (>15 bp), and the correction of overlapping paired reads. K-mer analysis was performed by khist.sh with 21 k-mers. Genome size and heterozygosity were estimated using GenomeScope v1.0.0 (Vurture et al. 2017) with the parameter "-m 1000."

A preliminary genome assembly was completed using Flye v2.8.1 (Kolmogorov et al. 2019) with a minimum overlap between reads of 1,000 bp. This was followed by two rounds of self-polishing steps. To improve assembly accuracy, two rounds of sequence polishing and redundant sequence removal were then carried out. Minimap2 v2.17 (Li 2018) was used to generate a sequence map.

NextPolish v1.1.0 (Hu et al. 2020) was employed to polish the preliminary assembly with Illumina short reads; this was performed twice. Redundant sequences were then removed using Purge_Dups v1.0.0 (Guan et al. 2020) with the parameter "-a 60."

Hi-C quality control and read alignment to the preliminary genome were performed using Juicer v.1.6.2 (Durand et al. 2016). The 3D de novo assembly (3D-DNA) v.180922 (Dudchenko et al. 2017) pipeline was used to correct misassemblies present and to conduct chromosome anchoring based on Hi-C sequences. Juicebox v1.11.08 (Durand et al. 2016) was used to manually correct possible errors from the first round of alignment by visualizing Hi-C heatmaps. The resulting assembly was then further refined by a second round of alignment using 3D-DNA. To remove potential contaminant sequences, we used BLAST+ (blastn) v2.7.1 (Camacho et al. 2009) to identify potential contaminant sequences by querying the NCBI nucleotide (nt) and UniVec databases.

We used BUSCO v3.0.2 (Waterhouse et al. 2018) to evaluate the completeness of the assembly with the parameter "-l insecta_odb10 -sp fly -m genome." Finally, Minimap2 v2.17 and SAMtools v.1.10 (Li 2009) were used to calculate the mapping rate by aligning PacBio long reads and Illumina short reads to the final chromosome-level assembly. Illumina reads of female adults were aligned to the chromosome-level assembly to identify the sex chromosome based on the average sequencing depth of *C. charites*.

Genome Annotation

We built a custom library to annotate repetitive elements using two different strategies: ab initio and homology searching. We generated a de novo repeat library for *C. charites* using RepeatModeler v2.0.1 (Flynn et al. 2020) based on repetitive element self-specificity with the parameter "-LTRStruct." We combined the de novo repeat library, Dfam v.3.3 (Storer et al. 2021), and the RepBase-20181026 (Bao et al. 2015) database to build a custom library. We masked the repetitive elements within the assembly with RepeatMasker v4.0.7 (Smit et al. 2013–2015) based on the above custom library. tRNAscan-SE v2.0.7 (Chan and Lowe 2019) was used to predict transfer RNAs and Infernal v1.1.2 (Nawrocki and Eddy 2013) was used to identify other remaining ncRNAs.

The PCGs were annotated using three distinct strategies: ab initio, homology-based, and transcriptome-based prediction. BRAKER v2.1.5 (Hoff et al. 2016) was used for ab initio gene prediction; this integrated Augustus v3.3.3 (Stanke et al. 2004) and GeneMark v4.32 (Brůna et al. 2020) concurrently with evidence from transcriptome and reference proteins to accurately model sequence properties. Transcriptome assembly alignment to the reference genome was performed using StringTie v2.1.4 (Kovaka et al. 2019). The transcriptome was provided as BAM alignments produced by HISAT2 v2.2.0 (Kim et al. 2015), and arthropod reference protein data were obtained from the OrthoDB v10.1 database (Kriventseva et al. 2019). The protein sequences of *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera litura*, *Helicoverpa armigera*, and *Tribolium castaneum* were downloaded from the NCBI database for the protein homology data set. GeMoMa (Keilwagen et al. 2019) was used for homology prediction. We used MAKER v3.01.03 (Holt and Yandell 2011) pipeline for genome annotation by integrating three different data sources listed above.

Diamond v2.0.8 (Buchfink et al. 2021) was used to search the UniProtKB database (Morgat et al. 2020) with the parameter "–more-sensitive -e 1e-5." Concurrently, eggNOG mapper v2.0 (Cantalapiedra et al. 2021) and InterProScan 5.48–83.0 (Jones et al. 2014) were used to annotate protein domains, GO designations, and KEGG pathways (KEGG, Reactome). eggNOG was used to search the eggNOG v5.0 (Huerta-Cepas et al. 2019) database, and InterProScan was used to query six databases, including Pfam, Panther, Gene3D, Superfamily, SMART, and CDD (Wilson et al. 2009; Marchler-Bauer et al. 2017; Letunic and Bork 2018; Lewis et al. 2018; El-Gebali et al. 2019; Mi et al. 2019). Finally, a diagram visualizing key genomic characteristics was drawn by TBtools v1.0692 (Chen et al. 2020).

Gene Family Evolution and Phylogeny

OrthoFinder v2.3.8 (Emms and Kelly 2019) was used to infer sequence orthology based on high-quality nonredundant protein sequences of 11 insect species downloaded from NCBI and GigaDB. These species included one trichopteran species (*Stenopsyche tienmushanensis*), three lepidopteran species (*B. mori*, *Danaus plexippus*, *H. armigera*), one dipteran (*D. melanogaster*), one coleopteran species (*T. castaneum*), one hymenopteran species (*Apis mellifera*), two paraneopteran species (*Rhopalosiphum maidis* and *Thrips palmi*), one isopteran species (*Coptotermes formosanus*), and one palaeopteran species (*Cloeon dipterum*; Supplementary table S3, Supplementary Material online).

Subsequently, MAFFT v7.450 (Katoh and Standley 2013) was used to align single-copy orthologous sequences via the L-INS-I method. TrimAl v1.4.1 (Capella-Gutierrez et al. 2009) was performed to trim alignments with the parameter "automated1." Next, FASconCAT-G v1.1.1 (Kück and Meusemann 2010) was used to concatenate them. IQ-TREE v2.07 (Minh et al. 2020) was used to determine which genes violated stationary, reversibility, and homogeneity assumptions using the parameters: "–symtest-pval 0.10"; it was then also used to construct a phylogenetic

tree using the general heterogeneous evolution (Ghost) strategy "-m LG + FO + H4." The posterior mean-site frequency model was used with the following set of parameters: "-m LG + C60 + FO + R" in IQ-TREE. The Ghost tree was used as an initial guide tree. MCMCTree within PAML v4.9j (Yang 2007) was used to estimate divergence time. Six fossils from the PBDB database (https://www.paleobiodb.org/navigator/) were used for node calibration, including the root: Pterygota (<443.4 Ma): Holometabola: 315.2–382.7 Ma (Bashkirian–late Devonian); Paraneoptera (Moscovian): 314.6–323.2 Ma; Coleoptera + Neuroptera (Moscovian): 307.0–323.2 Ma. CAFE v.5.0.0 (Mendes et al. 2021) was used to calculate the expansion or contraction of gene families based on final divergence time and the phylogeny with the default parameter.

Supplementary material

Supplementary material is available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

Data Availability

Genome assembly and raw sequencing data have been deposited at NCBI under the accessions numbers JACZEN000000000 and SRR13610846–SRR13610849, respectively. Genome annotations are available at 240 Figshare and can be accessed at https://figshare.com/projects/The_frist_chromosome-level_genome_assembly_of_Cheumatopsyche_charites_Malicky_Chantaramongkol_1997_Trichoptera_Hydropsychidae_to_understand_its_response_to_pollution/138072.

Literature Cited

Bao W, Kojima KK, Kohany O. 2015. Repbase update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 6(11):1–6.

Botosaneanu M, Szitó A, Oláh J. 1990. Macrozoobenthos communities in Hungarian lowland rivers. Aqua Hunga (Szarvas). 6:133–152.

Brůna T, Lomsadze A, Borodovsky M. 2020. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. NAR Genom Bioinform. 2(2):lqaa026.

Buchfink B, Reuter K, Drost HG. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods 18:366–368.

Bushnell B. 2014. BBtools. https://sourceforge.net/projects/bbmap/ (accessed 13 June 2019).

Camacho C, et al. 2009. BLAST+: architecture and applications. BMC Bioinf. 10(1):421.

Cantalapiedra CP, et al. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 38(12):5825–5829.

Capella-Gutierrez S, Silla-Martınez JM, Gabaldon T. 2009. Trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25(15):1972–1973.

Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences. Methods Mol Biol. 1962:1–14.

Chen C, et al. 2020. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant. 13(8):1194–1202.

Clair RMS, Dean JC, Flint OSJ. 2018. Description of adults and immature stages of *Antipodoecia* Mosely from Australia and synonymy of the families Antipodoeciidae and Anomalopsychidae (Insecta: Trichoptera). Zootaxa 4532(1):125–136.

Coffman WP, Cummins KW, Wuycheck JC. 1971. Energy flow in a woodland stream ecosystem. I. Tissue support trophic structure of the autumnal community. Archiv für Hydrobiol. 68(2):232–276.

Dudchenko O, et al. 2017. De novo assembly of the aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science 356-(6333):92–95.

Durand NC, et al. 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 3(1):95–98.

El-Gebali S, et al. 2019. The Pfam protein families database in 2019. Nucleic Acids Res. 47(D1):D427–D432.

Emms DM, Kelly S. 2019. Orthofinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20(1):238.

Flynn JM, et al. 2020. Repeatmodeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 117-(17):9451–9457.

Ge XY, Wang YC, Wang BX, Sun CH. 2020. Descriptions of larvae of three species of *Hydropsyche* Pictet 1834 (Trichoptera, Hydropsychidae) from China. Zootaxa 4858(3):4853–4858.

Geraci CJ, Kjer KM, Morse JC, Blahnik RJ. 2005. Phylogenetic relationships of Hydropsychidae subfamilies based on morphology and DNA sequence data. In: Tanida K, Rossiter A, editors. Proceedings of the 11thInternational Symposium on Trichoptera. Kanagawa: Tokai University Press. p. 131–136.

Gibbs DG. 1973. The Trichoptera of Ghana. Deut Entomol Z (NF). 20: 363–424.

Guan D, et al. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics 36(9):2896–2898.

Heckenhauer J, et al. 2019. Annotated draft genomes of two caddisfly species *Plectrocnemia conspersa* CURTIS and *Hydropsyche tenuis* NAVAS (Insecta: Trichoptera). Genome Biol Evol. 11(12): 3445–3451.

Heckenhauer J, et al. 2022. Genome size evolution in the diverse insect order Trichoptera. GigaScience 11:giac011.

Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics 32(5):767–769.

Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinform. 12(1):1–14.

Holzenthal RW, Blahnik RJ, Kjer KM, Prather AP. 2007. An update on the phylogeny of caddisflies (Trichoptera). In: Bueno-Soria J, Barba-Alvarez R, Armitage B, editors. Proceedings of the 12th International Symposium on Trichoptera. Columbus: The Caddis Press. p. 143–153.

Holzenthal RW, Morse JC, Kjer KM. 2011. Order trichoptera kirby, 1813. In: Zhang, Z.-Q. (Ed.) Animal biodiversity: an outline of

higher-level classification and survey of taxonomic richness. Zootaxa 3148(1):209–211.

Hu J, Fan J, Sun ZY, Liu SL, Berger B. 2020. Nextpolish: a fast and efficient genome polishing tool for long read assembly. Bioinformatics 36(7):2253–2255.

Huerta-Cepas J, et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 47(D1):D309–D314.

Jonason KA, Malm T, Espeland M. 2017. Molecular phylogeny of Sericostomatoidea (Trichoptera) with the establishment of three new families. Syst Entomol. 42(1):240–266.

Jones P, et al. 2014. Interproscan 5: genome-scale protein function classification. Bioinformatics 30(9):1236–1240.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.

Keilwagen J, Hartung F, Grau J. 2019. Gemoma: homology-based gene prediction utilizing intron position conservation and RNA-seq data. In: Kollmar M, editor. Gene prediction. New York (NY): Humana. p. 161–177.

Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 12(4):357–360.

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 37(5):540–546.

Kovaka S, et al. 2019. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. 20(1):1–13.

Kriventseva EV, et al. 2019. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Res. 47(D1):D807–D811.

Kück P, Meusemann K. 2010. FASconCAT: convenient handling of data matrices. Mol Phylogenet Evol. 56(3):1115–1118.

Letunic I, Bork P. 2018. 20 Years of the SMART protein domain annotation resource. Nucleic Acids Res. 46(D1):D493–D496.

Lewis TE, et al. 2018. Gene3D: extensive prediction of globular domains in proteins. Nucleic Acids Res. 46(D1):D435–D439.

Li H, et al. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25(16):2078–2079.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34(18):3094–3100.

Marchler-Bauer A, et al. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 45(D1):D200–D203.

Mendes FK, et al. 2021. CAFE 5 models variation in evolutionary rates among gene families. Bioinformatics 36(22–23):5516–5518.

Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. 2019. PANTHER Version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. 47-(D1):D419–D426.

Minh BQ, et al. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 37(5):1530–1534.

Misof B, et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science 346(6210):763–767.

Morgat A, et al. 2020. Enzyme annotation in UniProtKB using Rhea. Bioinformatics 36(6):1896–1901.

Morse JC. 2022. Trichoptera world checklist. https://entweb.sites.clemson.edu/database/trichopt/ (accessed 12 May 2022).

Morse JC, Frandsen PB, Graf W, Thomas JA. 2019. Diversity and ecosystem services of Trichoptera. Insects 10(5):125.

Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29(22):2933–2935.

Oláh J, Johanson KA, Barnard PC. 2008. Revision of the oriental and afrotropical species of Cheumatopsyche Wallengren (Hydropsychidae, Trichoptera). Auckland: Magnolia Press.

Smit AFA, Hubley R, Green P. 2013–2015. RepeatMasker Open-4.0. http://www.repeatmasker.org (accessed 19 May 2021).

Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. 32:W309–W312.

Storer J, Hubley R, Rosen J, Wheeler JR, Smit AR. 2021. The Dfam community resource of transposable element families, sequence models, and genome annotations. Mob DNA. 12(1):2.

Thomas JA, Frandsen PB, Prendini E, Zhou X, Holzenthal RW. 2020. A multigene phylogeny and timeline for Trichoptera (Insecta). Syst Entomol. 45(3):670–686.

Vurture GW, et al. 2017. Genomescope: fast reference-free genome profiling from short reads. Bioinformatics 33(14):2202–2204.

Wang YY, et al. 2022. The first chromosome-level genome assembly of a green lacewing Chrysopa pallens and its implication for biological control. Mol Ecol Resour. 22:755–767.

Waterhouse RM, et al. 2018. BUSCO Applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol. 35(3):543–548.

Wiggins GB. 1996. Larvae of the North American caddisfly genera (Trichoptera). 2nd ed. Toronto: University of Toronto Press.

Wilson D, et al. 2009. SUPERFAMILY–Sophisticated comparative genomics, data mining, visualization and phylogeny. Nucleic Acids Res. 37:D380–D386.

Wipfler B, et al. 2019. Evolutionary history of Polyneoptera and its implications for our understanding of early winged insects. Proc Natl Acad Sci U S A. 116(8):3024–3029.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586–1591.