Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Chromosome-level genome assembly of Chironomus striatipennis Kieffer provides insights into benthic adaptation and metamorphosis mechanism

Guo Bingxin[1]

[1]Shanghai Ocean University

July 6, 2022

Abstract

Chironomid is the most important macroinvertebrate species in aquatic ecosystem. Chironomus represents the most important genus of the Chironomidae. However, the species of this genus lack a high-quality assembled genome. Here, a high-quality chromosome-level assembled genome of Chironomus striatipennis which is an important model organism in aquatic ecological detection and toxicological application has been reported. The assembled genome size of C. striatipennis was 181.84 Mb, with a scaffold N50 value of 54.13 Mb. Furthermore, the molecular mechanism of adaptive evolution of Chironomid to benthic environment was elucidated by combining transcriptome data of different stages. The complete metabolic pathway of Hemoglobin was clarified in C. striatipennis for the first time to suggests the regulatory mechanism underlying its adaptation to benthic living. The expansions of CYP450s gene family related to detoxification explain its tolerance to the harsh environment. The key gene family, JHAMT, involved in biosynthesis of juvenile hormone are substantially expanded. The expansion of JHAMT genes and the regular regulation of juvenile hormone and ecdysone explain the developmental plasticity of C. striatipennis. In this study, it was also found that C. striatipennis is more dependent on JNK signal pathway induced metamorphosis than Drosophila melanogaster. This study provides some views into genetic basis of tolerance and adaptation of C. striatipennis to harsh benthic environments and lays a part of the foundation for the adaptive evolution of benthic animals.

Chromosome-level genome assembly of *Chironomus striatipennis* Kieffer provides insights into benthic adaptation and metamorphosis mechanism

Bingxin Guo[1], Jinxian Chen[1], Liang Lu[2], Li Wang[2], Liqing Wang[1], Chenhong Li[2], Wenbing Liu[3], Ruilei Zhang[1#]

[1]Enginnering Research Center of Environmental DNA and Ecological Water Health Assessment, College of Fisheries and life Science, Shanghai Ocean University, Shanghai 201306, China

[2]The Lab of Molecular Systematics and Ecology, College of Fisheries and Life Science, Shanghai Ocean University, Shanghai 201306, China

[3]Tianjin Key Laboratory of Conservation and Utilization of Animal Diversity, Tianjin Normal University, Tianjin, China

#Bingxin Guo and Jinxian Chen contributed equally to this work.

#To whom correspondence should be addressed. Email:*rlzhang@shou.edu.cn*

Abstract : Chironomid is the most important macroinvertebrate species in aquatic ecosystem. *Chironomus* represents the most important genus of the Chironomidae. However, the species of this genus lack a high-quality assembled genome. Here, a high-quality chromosome-level assembled genome of *Chironomus striatipennis* which is an important model organism in aquatic ecological detection and toxicological application

Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

has been reported. The assembled genome size of *C. striatipennis* was 181.84 Mb, with a scaffold N50 value of 54.13 Mb. Furthermore, the molecular mechanism of adaptive evolution of Chironomid to benthic environment was elucidated by combining transcriptome data of different stages. The complete metabolic pathway of Hemoglobin was clarified in C. striatipennis for the first time to suggests the regulatory mechanism underlying its adaptation to benthic living. The expansions of CYP450s gene family related to detoxification explain its tolerance to the harsh environment. The key gene family, JHAMT, involved in biosynthesis of juvenile hormone are substantially expanded. The expansion of JHAMT genes and the regular regulation of juvenile hormone and ecdysone explain the developmental plasticity of C. striatipennis. In this study, it was also found that *C. striatipennis* is more dependent on JNK signal pathway induced metamorphosis than *Drosophila melanogaster* . This study provides some views into genetic basis of tolerance and adaptation of *C. striatipennis* to harsh benthic environments and lays a part of the foundation for the adaptive evolution of benthic animals.

Keywords

Chironomus striatipennis , genome, benthic adaptation, metamorphosis mechanism

1 Introduction

In aquatic ecosystem, benthos can directly or indirectly affect the production of fish or waterfowl (Smith, 2014). In many aquatic ecosystems the number of chironomid species present may account for at least 50% of the total macroinvertebrate species recorded (Armitage et al., 1995) and serve as backbone components of aquatic benthic ecosystem.

Chironomids are the most ubiquitous and frequently the most abundant insects in all types of freshwater. Under certain conditions, such as at low levels of dissolved oxygen, larval chironomid may be the only insect present in benthic sediments (Armitage et al., 1995). Species richness, wide distribution and tolerance to adverse conditions make them be frequently selected as good indicator organisms to detect water environment and pollution in ecological studies (Goto et al., 2011; Hinton H E, 1951; Pinder, 1986; Sogame & Kikawada, 2017). Recent research released the flexibility of Hemoglobin in response to xenobiotics may play a critical role in adaptive evolution in chironomids (Ha & Choi, 2008). Therefore, the regulation of heme metabolic pathway related to hemoglobin activity is particularly important. Another important factor facilitating to adaptation for environment is that many detoxification enzyme genes are highly expressed in chironomid (Kozeretska et al., 2022). Futhermore, chironomids are holometabolous insects undergoing four distinct stages, egg, larva, pupa and adult, and the larval stage accounts for the vast majority of the lifecycle. Specifically, after laying eggs, they hatch into larva in 1-2 day for most Chironomids; after that, chironomid experiences a period of pelagic life before settling on substrates, living benthic; when the external environment and its physiological conditions are suitable, it metamorphoses and emerges into adults (Armitage et al., 1995). Therefore, chironomid must be able to adapt to different living environments and have a set of mechanisms that can regulate metamorphosis and development. However, neither its biological mechanism to cope with the complex external environment nor its regulatory mechanism in the process of metamorphosis and eclosion have been explored. So combing with life cycle characteristics of chironomid, we can focus on exploring adaptation of chironomid larva to benthic life and regulation mechanism of metamorphosis and eclosion.

Using high-quality genome as a "scalpel" will help to probe the genetic basis of environmental adaptation and metamorphosis in these species. In the last decade, with the popularization and application of high-throughput sequencing technology, the genomes of more species have been released. However, in family chironomidae (Diptera, Insecta), the genomes of only 8 species have been sequenced and assembled (Cornette et al., 2016; Kelley et al., 2014; Kutsenko et al., 2014; Sun et al., 2021; Vicoso & Bachtrog, 2015). In previous researches, the genus *Chironomus* has only scaffold level assembled genome (ncbi), the precision and accuracy of its genome may not rest content with the increasingly high standard of experimental research. Meanwihle, it has been proved that there are great differences between the genomes of *Chironomus* and other genus in Chironomidae (Cornette et al., 2015). Therefore, it is essential to obtain a high-quality genome of *Chironomus*

Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

species.

Chironomus striatipennis Kieffer (Diptera: Chironomidae) is widely distributed in Holarctic, Oriental and Neotropical regions; its larvae frequently present in oligotrophic to eutrophic freshwater, such as lakes, swamps, rivers, sewage ditches, fish ponds and rice fields; it is regarded as an alien species to America derived from Asia by unintentional human transport (Al-Shami et al., 2012; Amora et al., 2015; Martin, 2017) (Figure 1). *C. striatipennis* is widespread in China (Wang et al., 2020); we had ever found its larvae infected municipal water systems in several cities in South China. Given ease to reproduce, short life cycle and the larval stage accounts for most of life cycle, the species is amenable to being maintained as a laboratory colony for study of various aspects of its biology and toxicology test (Lacerda et al., 2014; Platzer-Schltz & Welsch, 1969; Platzer-Schultz, 1970; Wu et al., 2021; Zhang et al., 2019). It is necessary to understand the genetic mechanism of *C. striatipennis* for unraveling chironomid adaptability and tolerance (Neff et al., 2021; Zhang et al., 2020; Zhang et al., 2021).

In present research, Oxford Nanepore technology and Hi-C technology were applied to generate a chromosome-level genome assembly of *C. striatipennis* Kieffer. Further, combined with transcriptome, the genetic basis of adaptation to benthic low oxygen substrate were revealed in chironomid larvae, which will set a foundation for exploring the evolution direction of macrozoobenthos in the benthic environment.

2 Test organisms and methods

2.1 Test organisms

Chironomus striatipennis Kieffer were cultivated and maintained under optimal conditions in the water ecological restoration Laboratory of Shanghai Ocean University. Organisms were reared in a glass chamber containing fine inorganic sediment (<10 mm), under a 16-hour light to 6-hour dark cycle and constant aeration. The water temperature was maintained at 25 ± 1 . The yeast extract solution (CAS Number: 8013-01-2 from Macklin) was provided as food every day.

2.2 DNA preparation and sequencing

In order to estimate genome size and heterozygosity, 50 male adults were collected from the same colony, and high-quality genomic DNA was extracted by Sangon Biotech Ezup column animal genomic DNA purification kit. The extracted DNA was used to construct the paired end 150 (PE150) library which was sequenced by illumina sequencer. 20.62 Gb clean data were obtained by quality control, the total sequencing depth was about 120.75x, the content of GC was about 29.07%, the proportion of Q20 was more than 96.96% and the proportion of Q30 was more than 91.91%.

Another 50 male adults from the same colony were selected and delivered to the Biomarker Biotechnology Co., Ltd to construct long-read sequenced DNA libraries and sequence by the Oxford Nanepore Technologies (ONT). A total of 20.67 Gb raw data with a N50 read length of 24.7 kb were acquired through the third-generation nanopore platform sequencer. 20.32 Gb clean data were obtained after filtering adapter, short fragments and low-quality data.

Furthermore, another 500 male adults from the same colony were selected to construct the Hi-C DNA libraries which was sequenced through illumina sequencing platform; a total of 22.92 GB clean data was obtained with the proportion of Q30 reaching more than 91.46 %. After the library quality evaluation, the proportion of reads containing enzyme digestion sites in Hi-C library reached more than 31.09 % (Supplementary Table 1). The transcriptomes of 50 male adults of *C. striatipennis*Kieffer was sequenced from the same colony and the transcriptomes of*C. riparius* Meigen was downloaded from NCBI, both of them were used to assist genome annotation.

2.3 Genome draft assembly

Genome size and heterozygosity were measured using Jellyfish v.2.1.4 and GenomeScope v2.0 (Marcais & Kingsford, 2011; Vurture et al., 2017; Xiao et al., 2019); parameters setting were as follow -k 19 -p 2 -m 100000. The error of the nanopore clean reads was first corrected using NextDenovo

Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

(*https://github.com/Nextomics/NextDenovo)*, with the seed cutoff set at 28491bp. The genome draft of *C. striatipennis* was assembled also using NextDenovo. The nanopore assembled genome draft was polished in three runs for error-corrected long reads using the illumina DNA short reads by NextPolish v1.3.1 (Hu J. et al., 2020). After each polish, the completeness and quality of assembled genome draft of *C. striatipennis* were assessed using BUSCO based on the dipterta_odb10 database. The genome of the best BUSCO evaluation result was selected as genome draft for subsequent analysis (Manni et al., 2021).

2.4 Hi-C library construction and chromosome assembly

The raw reads produced by the illumina High-throughput sequencing platform based on the technology of sequencing by synthesis were filtered by SOAPnuke software (Chen et al., 2018). Juicer was employed to compare the obtained clean reads to genome draft (Durand et al., 2016). After filtering the results and removing the misaligned reads, 3d-DNA software was applied to preliminarily cluster, order, and orient the unique, valid interaction pairs onto the pseudochromosomes (Dudchenko et al., 2017). Juicer-box was used to correct pseudochromosomes to improve chromosome assembly level. For the evaluation of the Hi-C assembly results, the final pseudochromosome assemblies were divided into 100 kb bins of equal lengths and a heat map was obtained to visualize the interaction signals generated by the valid mapped read pairs between each bin.

2.5 Genome annotation

Repetitive sequences clustered or dispersed among genes widely exist in eukaryotic genomes. According to their distribution, the repetitive sequences are divided into interspersed repeats and tandem repeats. The repetitive sequences of *C. striatipennis* were annotated mainly by de novo prediction and homology-based searches. RepeatMasker v4.1.0 and RepeatProteinMask v4.1.0 softwares were employed to recognize repetitive sequences referring to the RepBaseRepeatMaskerEdition-2018.10.26 database (Bao et al., 2015). For de novo prediction, a repetitive sequence database was built using RepeatModeler v2.0.2a (Price et al., 2005), then the protein-coding sequences were filtered using the BLAST in the repetitive sequence database (Kent, 2002). RepeatMasker software was subsequently applied to predict repeat sequences in the genome. LTR_finder and LTR_ retriever were used to predict long terminal repeated (LTR) and the Tandem repeats search was carried out by the Tandem Repeat Finder (TRF) (Benson, 1999; Xu & Wang, 2007).

The annotation of high-quality protein-coding genes was realized by integrating homology-based, de novo and transcriptome-based predictions. For homology-based prediction, protein sequences of five species (*Anopheles gambiae* , *Culex quinquefasciatus* ,*Drosophil melanogaster* , *Musca domestica* ,*Polypedilum vanderplanki*) and the RNA-seq data of *C. striatipennis* which was assembled into full-length transcriptome through Trinity v2.8.5 software were merged to align and predict genome sequences via Maker v2.31.10 (Grabherr et al., 2011; Haas et al., 2013; Henschel et al., 2012; Holt & Yandell, 2011; Stanke et al., 2006). Then, the complete sequences of 15586 genes derived from homology prediction method were utilized to construct a hidden markov model through Augustus v3.4.0 and SNAP v2017-03-01. In this process, BUSCO was used to accelerate the model constructing of Augustus (Johnson et al., 2008; Manni et al., 2021; Stanke et al., 2006). Finally, Maker was applied to annotate and integrate the results generated by the above methods.

The predicted gene model was further functionally annotated by using protein database. Diamond was used to annotate the predicted protein coding genes by alignment to SwissProt (http://www.uniprot.org/) and NCBI Nr database (Buchfink et al., 2021; Buchfink et al., 2015). KAAS web server (https://www.genome.jp/kaas-bin/kaas_main) for KEGG annotations was accessed in March 2022. InterPro and GO (The Gene Ontology Consortium) databases were compared with the predicted protein coding genes by InterProScan(Jones et al., 2014).

For non-coding RNA annotations, tRNAscan-SE 2.0.9 was used to annotate tRNA sequences (Lowe & Eddy, 1997); Infernal v1.1.4 (*http://infernal.janelia.org/*) was used to predict others non-coding RNA through Rfam 14.7 (Nawrocki & Eddy, 2013).

2.6 Gene family analysis

Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

The protein sequences of 11 species in Diptera were selected for phylogenetic analysis, including 6 species of family Chironomidae (*C. striatipennis* , *Chironomus riparius, Chironomus tentans, Clunio marinus* , *Polypedilum vanderplanki, Propsilocerus akamusi*), 3 species of family Culicidae (*Anopheles gambiae, Anopheles sinensis, Culex quinquefasciatus*), one species of family Drosophilidae (*Drosophila melanogaster*) and one species of family Muscidae (*Musca domestica*) (Supplementary Table 4, 7). For further analysis, the script was used to extract the longest transcript of each gene; Orthofinder v2.2.6 was employed to identify gene family clusters (Emms & Kelly, 2015, 2019).

Multiple sequence alignment of single copy gene families generated from Orthofinder were performed to infer phylogeny of above 11 species by Mafft v7.407 (Rozewicki et al., 2019). Protein-aligned sequences were translated into coding sequences (CDS) and further optimized by Gblocks 0.91b (Castresana, 2000). The optimization results were connected into super gene and put into IQTREE v1.5.5 to construct phylogenetic tree (Nguyen et al., 2015). The divergence time was estimated by MCMCTREE in PAML package. The standard divergence time was obtained by Timetree (Yang, 1997). Based on the results of gene family clustering and phylogenetic tree, expansion and contraction of gene families were inferred. The significance of each expanded and contracted gene family was evaluated by CAFE v4.2 (De Bie et al., 2006). The KEGG annotation of gene families was performed using the same method as gene function annotation. Homologous gene pairs in the sequence were sought by BLAST (Kent, 2002). Colinear regions were recognized by McscanX (Wang et al., 2012).

2.7 Identification of gene family members

To understand the mechanisms of benthic adaptation and metamorphosis in Chironomidae, the key gene families associated with in the genomes of*C. striatipennis* , *C. riparius* , *C. tentans* , *C. marinus* , *P. vanderplanki* and *P. akamus* i were analyzed.

To identify Globin genes, all of 347 Globin sequences of Chironomidae were downloaded from NCBI Genbank and then BLASTP was used to search Homologous protein gene sequences (Kent, 2002). The Globin hmmer from Pfam (http://pfam.xfam.org) was employed to train new species Hidden Markov Model (HMM) by HMMER 3.0 (http://www.hmmer.org/) and search for genome (Jiang et al., 2008). The common genes were combined to obtain the putative gene families which were searched from the Pfam to confirm that it contained NBS conserved domain.

To identify CYP450s genes and GSTs genes, the GSTs sequences and the CYP450 reference protein sequences were downloaded from NCBI Genbank, respectively, and then BLASTP was used to search Homologous protein gene sequences (Kent, 2002). The GSTs homologous gene sequence and the P450 hmmer from Pfam (http://pfam.xfam.org) were used to train new species Hidden Markov Model (HMM) by HMMER 3.0 (http://www.hmmer.org/) and search for genome, respectively (Jiang et al., 2008). The common genes were combined to obtain the putative gene families which were searched from the Pfam to confirm that it contained NBS conserved domain.

To identify JHAMT(juvenile hormone acid methyl transferase)genes and CASP7 genes,all of the JHAMT sequences and the CASP 7 sequences of*Drosophila* were downloaded from NCBI Genbank, and then BLASTP was used to search Homologous protein gene sequences, respectively (Kent, 2002). The Homologous protein gene sequences were used to train the JHAMT hmm model and CASP7 hmm model, respectively. The newly trained hmm model was used to search their gene sequences (Jiang et al., 2008). Finally, we combined the common genes as the putative gene families.

The protein sequences of each gene families were aligned by Muscle v3.8.1551 and phylogenetic tree was constructed by IQTREE with default parameters (Katoh & Standley, 2013). The phylogenetic tree was visualized using EvolView v3 (*https://www.evolgenius.info/*) (Subramanian et al., 2019).

2.8 Identification of genes involved in the metabolism pathway in *C. striatipennis*

The annotated proteins of *C. striatipennis* were mapped into KEGG's pathway map to indentify the sequence of all 13 genes or genes family members involved in tricarboxylic acid cycle pathway. And then

Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

the required genes from the pathway were deducted. Finally, the extracted protein was verified by KEGG annotation (Kanehisa et al., 2004). Genes and genes families related to porphyrin metabolism, development and metamorphosis were searched using the same method.

2.9 **Gene expression analysis**

Total RNA of *C. striatipennis* was respectively extracted from larval, pupal and adult individuals. For each library, three independent replicate RNA samples were prepared. The RNA samples were then sequenced on the illumina platform by biomarker Biotechnology Co., Ltd (Supplementary Table 14). After removing the adapter primers and low-quality reads, the clean reads were obtained for follow-up analysis. The clean reads were mapped to reference genomes by HISAT2 (Kim et al., 2015). The reads mapped on the genome were assembled into complete transcripts by StringTie to evaluate their expression level (Pertea et al., 2015). Differential expression gene set (DEG set) was obtained by differential expression analysis between sample groups using Deseq2 (Love et al., 2014; Trapnell et al., 2012). Then, the functions of differentially expressed genes were annotated based on the Go and KEGG database.

3 Results and Discussion

3.1 Species identification

C. striatipennis Kieffer had ever been confusedly identified as*C. kiiensis* Tokunaga or *C. strenzkei* Fittkau (Lacerda et al., 2014). It widely distributes in China and was always under the name of *C. kiiensis* Tokunaga. In present study, the divergences of the cytochrome oxidase subunit 1 (COI) gene between this study and Martin's released were 0.46% by quantifying the genetic distance matrix (Supplementary Table 2). This genetic distance divergence is within the range of genetic distance among most species (Hebert et al., 2003). Therefore, the laboratory colony of chironomid used in present study is inseparable from *C . striatipennis* named by Martin on morphology and DNA barcodes, so we follow Martin's identification and name it *C. striatipennis* Kieffer (Amora et al., 2015).

3.2**Genome sequencing and characteristics**

The estimated genome size of *C. striatipennis* was about 170.79 Mb, the heterozygosity rate was about 1.13%, the repeat sequence part of genome was about 23.63% and with karyotype of 2N=2X=8, as determined through K-mer analysis (Supplementary Figure 1). These characteristies indicated that *C. striatipennis* has a highly heterozygous complex genome.

In order to obtain a better assembly, Oxford Nanepore Technologies sequencing data was used to the preliminary genome draft assembly and then using the illumine data to polish the preliminary genome draft. BUSCO assessment indicated that the completeness of the gene set of assembled genome draft was 95.0%, which signified the genome assembly of *C. striatipennis* was complete and suit for further anchoring sequences to chromosome analysis (Supplementary Figure 2, Supplementary Table 3). Basing on 22.92 Gb clean reads from Hi-C library, 78 scaffolds including 4 pseudomolecules which represented 4 chromosomes and 74 detritus were assembled. The lengths of 4 pseudochromosomes ranged from 20.40 Mb to 60.13 Mb with a scaffold N50 value of 64.51 Mb (Figure 3). In addition, about 179.77 Mb contigs were mapped into 4 pseudochromosomes with an anchoring rate of 98.86 % (Supplementary Table 5). The chromatin interaction data suggest that our Hi-C assembly is of high quality (Figure 2). We used BUSCO to identify 95.0% (3119/3285), 98.7% (1349/1367) and 98.1% (936/954) conserved genes of*C. striatipennis* by alignment to corresponding database of Diptera, Insecta and Metazoa (Supplementary Figure 3). The above results are compared with other genomes assembled in Chironomid, it can be concluded that the genome assembly of *C. striatipennis* was more high-quality and complete (Tab.1).

C. striatipennis is the first species with chromosome-level genome assembly in genus *Chironomus* . The genome size of *C. striatipennis* is similar to other species in the genus*Chironomus* (C. riparius, C.tentans, C. tepperi(236 Mb)), but it is much larger than that of species in subfamily Orthocladiinae (*C. marinus* , *Belgica antarctica* (99 Mb) and *P. akamusi*) and in genus *Polypedilum* (*P. vanderplanki* ,*P. pembai* (122 Mb)) (Tab.1) (Kaiser et al., 2016; Kelley et al., 2014; Sun et al., 2021). The above results are also consistent

Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

with those of Cronette et al. (2015). In conclusion, it can be inferred the genome size of genus *Chironomus* is larger than that of subfamily Orthocladiinae and genus *Polypedilum* .

Table 1. Genome statistics and comparisons among chironomid species whose genome has been sequenced

3.3 Gene prediction and annotation

Total 13429 protein-coding genes were obtained through de novo and homologous prediction. The genome assembly completeness assessed by BUSCO was up to 98.7% (Tab 1), which means the genome annotation for*C. striatipennis* is reasonably complete.

Combined with the results of other studies, the length of gene and the average length of introns in *C. striatipennis* like other species in subfamily Chironominae were much higher than those of subfamily Orthocladinae (Supplementary Table 8)(Kutsenko et al., 2014; Vicoso & Bachtrog, 2015). The Protein coding part of the genome in C. striatipennis was smaller than those in other subfamily Orthocladinae (Supplementary Table 8). It is worth noting that the introns of *C. striatipennis* was concentrated at the 5 un-translated regions (5' UTR). These introns usually contain transcriptional regulatory elements and they can reduce gene mutations (Rethmeier et al., 1997; Rose, 2004; Rose & Beliakoff, 2000). In the future, the role of introns in regulating gene expression and reducing gene mutations should be further explored. Besides this, 105 ribosomal RNAs, 31 small nuclear RNAs, 34 microRNAs and 203 transfer RNAs were identified in *C. striatipennis*(Supplementary Table 9).

The repeated sequences comprised 23.29 % of whole genome in *C. striatipennis* which was much higher than other known chironomid genomes (Supplementary Table 8). Interspersed repeat is up to 78.3% and makes up a key component of repeated sequence. Moreover, the number of repeated sequences predicted by de novo was far more than that obtained based on RepBaseMaskerEdition-2018.10.26 database. It means that*C. striatipennis* has many special repeats sequence. The content of long terminal repeated (LTR) predicted by de novo makes up 21.14% of the repeat sequences of *C. striatipennis* (Supplementary Table 10-11)(Cornette et al., 2017; Kaiser et al., 2016; Kelley et al., 2014; Kutsenko et al., 2014; Shaikhutdinov & Gusev, 2022; Sun et al., 2021; Vicoso & Bachtrog, 2015).

A total of 12717 (94.70%) predicted genes were functionally annotated, among which 12245 (91.18%) genes were annotated with diamond based on Nr and 8292 (61.75%) genes were annotated with diamond based on SwissProt databases. 5579 (41.54%) genes were annotated through Kaas. The number of annotated genes searched by InterProScan based on Pfam, Go and InterPro databases were 1369 (10.19%), 3079 (22.93%) and 12135 (90.36%), respectively (Supplementary Table 12).

3.4 Genome evolution

2446 single-copy gene families shared by *C. striatipennis* and other 10 species in Diptera were used to construct phylogenetic tree and estimate divergence time (Supplementary Figure 4), which species-specific orthogroups were identified in a Venn diagram (Figure 4-b, Supplementary Table 6). Phylogenetic analysis shows that *C. striatipennis* is closely related to *C. riparius* and *C. tentans* , they form a clade located at the base of Chironomidae. The Chironomidae divergence was estimated to have occurred about 152.9 Mya, Orthocladiinae diverged from Chironominae about 71.7 Mya, the divergence of *C. striatipennis* occurred about 13.7 ˜ 43.6 Mya (Figure 4-a). This indicates that *Chironomus* belongs to a relatively young group in Chironomidae.

The analysis of expansion and contraction of orthologoups gene families show that the clad of Chironomidae had 28 gene families expanded and 18 gene families contracted. 105 gene families in *C. striatipennis*were expanded, by contrast, 256 in *C. rapirius* , 99 in *C. tentans* , 247 in *P. vanderplanki* , 53 in *C. marinus* and 25 in *P. akamusi* . Meanwhile, 70 gene families in *C. striatipennis* were contracted, by contrast, 80 in *C. rapirius* , 185 in *C. tentans* , 84 in *C. marinus* and 14 in *P. akamusi* (Supplementary Figure 5). In *C. striatipennis* , the expanded gene families were significantly involved in growth, development and defensive metabolism such as larva molting, damage repair and inflammatory immunity (Supplementary Figure 6), while the contracted gene families were mainly involved in energy metabolism (Supplementary Figure 7).

Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

These findings suggest that the expanded and contracted gene families may be closely related to the adaptive evolution in *C. striatipennis.* It is similar to *Belgica antarctica* and *P. akamusi* , that they can resist the adverse environment by expanded heat shock proteins (HSPs) (Sun, X et al., 2021; Kozeretska, I et al., 2022). Chironomid midge with same life history as*C. striatipennis* may also expand similar gene families to adapt to the external environment (Cornette et al., 2017; Kaiser et al., 2016; Kelley et al., 2014; Kutsenko et al., 2014; Shaikhutdinov & Gusev, 2022; Sun et al., 2021; Vicoso & Bachtrog, 2015).

In order to recognize homologous proteins sequences in the genome, BLASTP and McscanX were employed to analyze the protein sequences of*C. striatipennis* , *P. akamusi* and *P. vanderplanki* to obtain their collinear gene pairs. There were 7162 collinear gene pairs between *C. striatipennis* and *P. akamusi* , 12438 collinear gene pairs between *C. striatipennis* and *P. vanderplanki* , 270 collinear gene pairs between *C. striatipennis* and itself (Supplementary Table 13). Like flies, Chironomid midge also has very low self-collinearity at the chromosome level. However, there is a close correspondence between several chironomid species with chromosome level genome assembly. The possible reason for that is Chironomid midges maybe abandon some unnecessary genes in the process of evolution.

3.5 Gene families associated with adaptive evolution

Some key gene families, including the Globin related to oxygen transport, the JHAMT and caspase 7 related to regulating development and metamorphosis, the GSTs and CYP450s related to detoxification were identified to illuminate the adaptation of Chironomid larva to benthic environment and metamorphosis mechanism.

Globin for benthic adaptation. It is known to all, most invertebrates have copper containing Hemocyanin and non heme iron Hemoglobin. Except for a few annelids, the Chironomid midge is one of the few invertebrate with Hemoglobin (Strand et al., 2004). Hemoglobin can provide necessary dissolved oxygen for metabolic activities of organisms (Bergtrom et al., 1976). Most chironomid larvae live in aquatic substrate with low dissolved oxygen and are urgent for hemoglobin to meet oxygen deficiency. Globins are the key dissoxygen carriers after binding to heme. The evolution of Globin in chironomid midges can be inferred by sequencing the genome of *C. striatipennis* . We identified were 10 Globin genes of *C. striatipennis* . The 10 Globin can be divided into three subfamilies based on ncbi database, seven in the subfamily Globin III, two in the subfamily Globin VII and the other one in the subfamily Globin X (Figure 5-a). It should be noted that eight Globin genes belonging to three gene tandem clusters in the genome have different gene sequences and number of conserved functional regions, which indicates that the gene has experienced some convergent or divergent mutation after tandem expansion.

Considering in early larval stages, *C. striatipennis* shifts from pelagic to benthic habitat in water and then emerges on water surface, in order to verify the hemoglobin produced by porphyrin pathway to adapt to the situation of low dissolved oxygen in benthic habitat at larval stage, 8 genes or gene family members involved in porphyrin metabolic pathway were analyzed in whole life cycle. These genes include 5-aminolevulinate synthase (ALAS), porphobilinogen synthase (ALAD), hydroxymethylbilane synthase (HMBS), uroporphyrinogen III decarboxylase (UROS), uroporphyrinogen decarboxylase (UROD), coproporphyrinogen III oxidase (CPOX), protoporphyrinogen III oxidase (PPOX) and protoporphyrin ferrochelatase (FECH) (Frankenberg et al., 2001; Saffarini et al., 1991). These genes expressed significantly in larval benthic stage and began to close in pupal stage (Figure 7, Supplementary Table15,17). In addition, in *C. striatipennis* , among the 10 putative Globin gene family members, 8 genes which average fragments per kilobase of transcript per million mapped (FKPM) reads for Globin genes were 6.74-35492.96 were highly expressed in larval stage, but the expression level of most Globin genes decreased to less than 1% in pupal stage and 2 pseudogenes were not expressed in the whole life cycle (Figure 5-b, Supplementary Table 16).

These data provided convincing demonstration for the correlation between the differential expression of hemoglobin and the adaptation of chironomid larva to benthic life, which indicates the chironomid species with the same lifestyle as *C. striatipennis* have a highly effective oxygen transport system to adapt to the low dissolved oxygen in benthic habitat.

Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Detoxification. The insect can secrete a variety of enzymes by detoxification system to resist the influence of external poisons (Li et al., 2018). Generally, the detoxification system usually goes through three phases to resist external poisons (Epel et al., 2008; Martinez-Paz, 2018; Smital et al., 2004). However, no matter which phase is involved, Cytochrome P450s (CYP450s) and glutathione S-transferases (GSTs) enzymes play core roles in process of resisting the influence of external poisons. 37 CYP450s and 26 GSTs were identified in *C. straitipennis* in this study (Supplementary Figure 8). CYP450s not only play an importance role in detoxification metabolism and synthesis of hormones but also have an obvious phenomenon of gene family expansion. The phylogenetic analysis of the CYP450s in *C. striatipennis* showed they were divided into four branches: the CYP2, the CYP3, the CYP4 and the mitochondrial (Mito) (Figure 6-b). It is worth noting that most genes of the CYP450s in*C. striatipennis* are clustered under the branches of the CYP3 (15/37 genes) and the CYP4 (13/37 genes). The genes involved in these two gene subfamilies are often induced to express by exogenous toxic substances (David et al., 2010; Ffrench-Constant, 2013). More interestingly, *P. vanderplanki* is known as a "trisomic" creature in biological world that can survive in various extreme environments, and the number of CYP450s genes has been amplified to as high as 154 (Figure 6-a). Aboving results provide solid evidence that in chironomid midge CYP450s, especially the CYP3 and the CYP4 gene subfamily, are related to adapt to harsh benthic environment, resist interference of various toxic substances and survive successfully in some extreme environment (Lu et al., 2021).

Development. Chironomid midge has relative developmental plasticity. It can change the developmental trajectory to deal with different environments, which is also a strategy for adapting the environment. Like most insects, chironomid midge is also coordinately regulated by ecdysone and juvenile hormone from larva to pupa and to adult, juvenile hormone can prevent metamorphosis induced by ecdysone (Guo et al., 2019; Hu X. L. et al., 2019; Miki et al., 2020; Santos et al., 2019). The genes involved in biosynthesis of ecdysone and juvenile hormone which regulate insect development were also recognized in*C. striatipennis.* Notably, the key genes, JHAMT, involved in biosynthesis of juvenile hormone are substantially expanded. *C. striatipennis* harbors 43 JHAMT genes; the number is similar to *P. vanderplanki* (38 JHAMT) and much higher than the equivalent in other species in Chironomidae (Supplementary Figure 10).

It has been reported that there were very few chironomid midges can grow rapidly in the cold winter (Armitage et al., 1995). Most chironomid midges overwinter at the fourth larval instar and will not emerge into adults until the beginning of the next spring (Armitage et al., 1995). By foregoing reason, it is speculated that the expansion of JHAMT genes is a general strategy for chironomid to regulate the content of juvenile hormone to overwinter. Therefore, when the external situation is appropriate, the expression of JHAMT genes will decrease.

The transcriptomes were sequenced to reveal how chironomid regulates the synthesis of ecdysone and juvenile hormone during the process of molting, metamorphosis and growth in *C. striatipennis* . In present study, the expression level of various enzymes related to the biosynthesis of juvenile hormone decreased during the shift from larva to adult in *C. striatipennis* ; it is consistent with the hypothesis that the juvenile hormone of chironomid midge decreases before eclosion to ensure a smooth eclosion (Figure 7). Interestingly, the expression level of JHAMT enzyme closely related to the synthesis of juvenile hormone III slightly increased during the shift from pupa to adult, which may be due to chironomid midge needs to provide the time required for morphology and organelle remodeling during metamorphosis (Figure 8, Supplementary Table 18).

In addition, the expression level of ecdysone 20-monooxygenase involved in the regulation of ecdysone decreased during the period of eclosion. In conclusion, *C. striatipennis* controls the emergence time by flexibly adjusting the concentration of juvenile hormone, which displays development plasticity in chironomid midge.

Metamorphosis mechanism. Programmed cell death plays an important role in removing cells that are unnecessary or potentially dangerous to organisms, and apoptosis is one of the most important ways of programmed cell death (Baehrecke, 2002; Fuchs & Steller, 2015; Galluzzi et al., 2018; Hay et al., 2004). Apoptosis is essential not only to maintain homeostasis and immune response but also to its metamorphosis. Apoptosis can reshape the whole morphology of organisms for metamorphosis development to help them

Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

adapt to the new environment in the future. Previous studies have shown that ecdysone can regulate the metamorphosis of *D. melanogaster* by inducing the expression of apoptotic effectors. The specific molecular mechanism is that ecdysone activates the expression activity of inducing downstream apoptosis factor by binding to ECR/USP heterodimer, which initiates the caspase pathway of apoptosis by inhibiting the expression activity of apoptosis inhibiting factors (Hay & Guo, 2006; Hay et al., 2004). In present study, some key components involved in these pathways were found in *C. striatipennis* . Caspase 7, the effector of apoptosis, was significantly expanded in the genome of *C. striatipennis*(Supplementary Figure 9). Furthermore, the JNK-dependent apoptosis pathway is more active in *C. striatipennis* , which is not like*D. melanogaster* heavily relying on ecdysone to open the caspase pathway.

JNK pathway is an important signaling pathway involved in many life processes such as cell morphology construction, cell polarity establishment and apoptosis in cells (Kuranaga et al., 2002). In Diptera, the main role of JNK pathway is to promote thorax closure by assisting actin to pull bilateral thorax to the midline through apoptosis unnecessary tissue cells (Marti'n-Blanco et al., 2000; Zeitlinger & Bohmann, 1999). In this study, the expression level of components related to regulating JNK pathway ascended gradually with the advancement of metamorphosis in *C. striatipennis* (Figure 7, Supplementary Table 19). This means that wing shaping is the most important event in the process of *C. striatipennis* metamorphosis. In addition, similar as other arthropods, *C. striatipennis*constructs the first line of defense through the chitin shell, which can help chironomid avoid virus or various mechanical damages. The thorax closure can promote a pair of wings to move towards the midline and close to each other. Limited by the size of the individual, the original chitin pupal cortex was inevitably partially degraded in the process of thorax closure. Therefore, the gene encoding chitinase in the genome of*C. striatipennis* was also significantly expanded (Supplementary Figure 11). Interestingly, these chitinases have relative cascade effect in expression process, which was, some chitinases were extremely expressed during the transition from larva to pupa, while others were actively expressed during the shift from pupa to adult (Supplementary Table 20). These proteins work together in series with several signal pathways to promote the successful completion metamorphosis and development in *C. striatipennis* .

4. Conclusion

Here, a high-quality chromosome-level genome assembly of *C. striatipennis* with genome size of 181.84Mb and 13429 annotated protein-coding genes has been reported. Through phylogenetic analysis, the possible position of *C. striatipennis* in evolutionary relationship was clarified. By the analysis of expanded gene family, their possible role (such as hemoglobin, GST, CYP450s, JHAMT, etc.) in the adaptability of benthic environment were clarified. In addition, in the process of metamorphosis, *C. striatipennis* depends more on JNK signal pathway to initiate the apoptosis pathway, and then smoothly complete the eclosion process. The results in this research provide a relatively high-quality genome and a foundation for revealing the evolutionary direction of organisms in adverse environments.

Acknowledgments

The authors are grateful to financial support from Science and Technology Commission of Shanghai Municipality and National Natural Science Foundation of China. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Accessibility

The raw genomic data of Oxford Nanopore Technologies reported in this paper have been deposited in the NCBI database under Accession no. SRR19836228. The raw genomic data of illumina reported in this paper have been deposited in the the NCBI database under Accession no. SRR19843066. The raw genomic data of Hi-C reported in this paper have been deposited in the NCBI database under Accession no. SRR19843067. The*C. striatipennis* transcriptome data for genome annotation were submitted to NCBI under Accession no. SRR19843063. The Whole Genome Shotgun project has been deposited at GenBank (NCBI Assembly PRJNA852386). The transcriptomic data of *C. striatipennis* for different stages are available through Accession no. SRR19843054-62. The DNA barcode of cytochrome c oxidase I(COI)reported in this

Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

paper was uploaded to NCBI, and the accession number is ON878180-81.

ORCID

*Bingxin Guohttps://orcid.*org/0000-0002-7606-9560

Reference

Al-Shami, S. A., Rawi, C. S. M., Ahmad, A. H., & Nor, S. A. M. (2012). Redescription of Chironomus javanus and Chironomus kiiensis (Diptera: Chironomidae) Larvae and Adults Collected from a Rice Field in Pulau Pinang, Malaysia. *Tropical life sciences research, 23* (1), 77-86.

Amora, G., Hamada, N., Fusari, L. M., & Andrade-Souza, V. (2015). An Asiatic Chironomid in Brazil: morphology, DNA barcode and bionomics.*Zookeys* (514), 129-144. https://doi.org/10.3897/zookeys.514.9925

Armitage, P. D., Cranston, P. s., & Pinder, L. C. (1995). The Chironomidae: Biology and ecology of non-biting midges. *Springer Science: Chapman & Hall.* https://doi.org/10.1007/978-94-011-0715-0.

Baehrecke, E. H. (2002). How death shapes life during development.*Nat Rev Mol Cell Biol, 3* (10), 779-787. https://doi.org/10.1038/nrm931

Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA, 6* , 11. https://doi.org/10.1186/s13100-015-0041-9

Bellex, X., Martin D., Piulachs M. (2005). The Mevalonata Pathway and the synthesis of Juvenile hormone in insects. *Annu Rev Entomol* , 2005,50:181-99. https:/ /doi.org/10.1146/annurev.ento.50.071803.130356

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic acids research, 27* (2), 573-580. https://doi.org/ 10.1093/nar/27.2.573

Bergtrom, G., Laufer, H., & Rogers, R. (1976). Fat Body: A Site Of Hemoglobin Syntheis In Chironomus Thummi (Diptera) *Cell Biology, 69* , 264-274. https://doi.org/10.1083/jcb.69.2.264

Buchfink, B., Reuter, K., & Drost, H. G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods, 18* (4), 366-368. https://doi.org/10.1038/s41592-021-01101-x

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat Methods, 12* (1), 59-60. https://doi.org/10.1038/nmeth.3176

Castresana, J. (2000). Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution, 17* (4), 540-552. https://doi.org/10.1093/ oxfordjournals.molbev.a026334

Chen, Y., Chen, Y., Shi, C., Huang, Z., Zhang, Y., Li, S., Li, Y., Ye, J., Yu, C., Li, Z., Zhang, X., Wang, J., Yang, H., Fang, L., Chen, Q. (2018). SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience, 7* (1), 1-6. https://doi.org/10.1093/gigascience/gix120

Cornette, R., Gusev, O., Nakahara, Y., Shimura, S., Kikawada, T., & Okuda, T. (2015). Chironomid midges (Diptera, Chironomidae) show extremely small genome sizes. *Zoological Science* , 32(3), 248– 254. https:/ /doi.org/10.2108/zs140166.

Cornette, R., Yamamoto, N., Masaru, & Yamamoto. (2017). A new anhydrobiotic midge from Malawi, Polypedilum pembai sp.n. (Diptera: Chironomidae), closely related to the desiccation tolerant midge, Polypedilum vanderplanki Hinton. *Systematic Entomology, 42* , 814-825. https://doi.org/ 10.1111/syen.12248

David, J. P., Coissac, E., Melodelima, C., Poupardin, R., Riaz, M. A., Chandor-Proust, A., & Reynaud, S. (2010). Transcriptome response to pollutants and insecticides in the dengue vector Aedes aegypti using

Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

next-generation sequencing technology. *BMC Genomics, 11* , 216. https://doi.org/10.1186/1471-2164-11-216

De Bie, T., Cristianini, N., Demuth, J. P., & Hahn, M. W. (2006). CAFE: a computational tool for the study of gene family evolution.*Bioinformatics, 22* (10), 1269-1271. https://doi.org/10.1093/bioinformatics/btl097

Demay, Y., Perochon, J., Szuplewski, S., Mignotte, B., Gaumer, S. (2014). The PERK pathway independently triggers apoptosis and a Rac1/Slpr/JNK/Dilp8 signaling favoring tissue homeostasis in a chronic ER stress Drosophila model. Cell Death and Disease, 5, e1452. https://doi.org/10.1038/cddis.2014.403

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., Shamim, M. S., Machol, I., Lander, E. S., Aiden, E. L. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science, 356* , 92-97. https://doi.org/10.1126/science.aal3327

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S., Huntley, M. H., Lander, E. S., & Aiden, E. L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst, 3* (1), 95-98. https://doi.org/10.1016/j.cels.2016.07.002

Emms, D. M., & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol, 16* , 157. https://doi.org/10.1186/s13059-015-0721-2

Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol, 20* (1), 238. https://doi.org/10.1186/s13059-019-1832-y

Epel, D., Luckenbach, T., & Stevenson, C. N. (2008). EFFLUX TRANSPORTERS: Newly Appreciated Roles in Protection against Pollutants.*Environmental science & technology, 42* (11), 3914-3920. https://doi.org/10.1021/es087187v

Ffrench-Constant, R. H. (2013). The molecular genetics of insecticide resistance. *Genetics, 194* (4), 807-815. https://doi.org/10.1534/genetics.112.141895

Frankenberg, N., Mukougawa, K., Takayuki Kohchi, & Lagarias, J. C. (2001). Functional Genomic Analysis of the HY2 Family of Ferredoxin Dependent Bilin Reductases from Oxygenic Photosynthetic Organisms. *The Plant cell, 13* , 965–978. https://doi.org/ 10.2307/3871353

Fuchs, Y., & Steller, H. (2015). Live to die another way: modes of programmed cell death and the signals emanating from dying cells.*Nat Rev Mol Cell Biol, 16* (6), 329-344. https://doi.org/10.1038/nrm3999

Galluzzi, L., Vitale, I., Aaronson, S. A., Abrams, J. M., Adam, D., Agostinis, P., . . . Kroemer, G. (2018). Molecular mechanisms of cell death: recommendations of the Nomenclature Committee on Cell Death 2018.*Cell Death Differ, 25* (3), 486-541. https://doi.org/10.1038/s41418-017-0012-4

Goto, S. G., Philip, B. N., Teets, N. M., Kawarasaki, Y., Lee, R. E., Jr., & Denlinger, D. L. (2011). Functional characterization of an aquaporin in the Antarctic midge Belgica antarctica. *J Insect Physiol, 57* (8), 1106-1114. https://doi.org/10.1016/j.jinsphys.2011.03.023

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Regev, A. (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol, 29* (7), 644-652. https://doi.org/10.1038/nbt.1883

Guo, S., Sun, D., Tian, Z., Liu, W., Zhu, F., & Wang, X. P. (2019). The limited regulatory roles of juvenile hormone degradation pathways in reproductive diapause preparation of the cabbage beetle, Colaphellus bowringi. *J Insect Physiol, 119* , 103967. https://doi.org/10.1016/j.jinsphys.2019.103967

Ha, M. H., & Choi, J. (2008). Chemical-induced alteration of hemoglobin expression in the 4th instar larvae of Chironomus tentans Mg. (Diptera: Chironomidae). *Environ Toxicol Pharmacol, 25* (3), 393-398. https://doi.org/10.1016/j.etap.2007.12.006

Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., . . . Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc, 8* (8), 1494-1512. https://doi.org/10.1038/nprot.2013.084

Hay, B. A., & Guo, M. (2006). Caspase-dependent cell death in Drosophila. *Annu Rev Cell Dev Biol, 22* , 623-650. https://doi.org/10.1146/annurev.cellbio.21.012804.093845

Hay, B. A., Huh, J. R., & Guo, M. (2004). The genetics of cell death: approaches, insights and opportunities in Drosophila. *Nat Rev Genet, 5* (12), 911-922. https://doi.org/10.1038/nrg1491

Hebert, P. D., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proc Biol Sci, 270* (1512), 313-321. https://doi.org/10.1098/rspb.2002.2218

Henschel, R., Lieber, M., Wu, L.-S., Nista, P. M., & Haas, B. J. (2012). Trinity RNA-Seq Assembler Performance Optimization. *ACM, 1* . https://doi.org/10.1145/2335755.2335842

Hinton H.E. (1951). A new Chironomid from Africa, the larva of which can be dehydrated without injury. *Proceedings of the Zoological Society of London, 121* (2), 371-380. https://doi.org/ 10.1093/gigascience/gix120

Holt, C., & Yandell, M. (2011). MAKER2: an annotation pipeline and genome_database management tool for second_generation genome projects.*Bmc Bioinformatics, 12* , 491. https://doi.org/10.1186/1471-2105-12-491

Hu, J., Fan, J., Sun, Z., & Liu, S. (2020). NextPolish: a fast and efficient genome polishing tool for long-read assembly.*Bioinformatics, 36* (7), 2253-2255. https://doi.org/10.1093/bioinformatics/btz891

Hu, X. L., Niu, J. J., Meng, Q., Chai, Y. H., Chu, K. H., & Chan, K. M. (2019). Effects of two juvenile hormone analogue insecticides, fenoxycarb and methoprene, on Neocaridina davidi. *Environ Pollut, 253* , 89-99. https://doi.org/10.1016/j.envpol.2019.06.120

Jiang, K., Thorsen, O., Peters, A., Smith, B., & Sosa, C. P. (2008). An Efficient Parallel Implementation of the Hidden Markov Methods for Genomic Sequence-Search on a Massively Parallel System. *IEEE Transactions on Parallel and Distributed Systems, 19* (1), 15-23. https://doi.org/10.1109/tpds.2007.70712

Johnson, A. D., Handsaker, R. E., Pulit, S. L., Nizzari, M. M., O'Donnell, C. J., & de Bakker, P. I. (2008). SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap.*Bioinformatics, 24* (24), 2938-2939. https://doi.org/10.1093/bioinformatics/btn564

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., . . . Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics, 30* (9), 1236-1240. https://doi.org/10.1093/bioinformatics/btu031

Kaiser, T. S., Poehn, B., Szkiba, D., Preussner, M., Sedlazeck, F. J., Zrim, A., . . . Tessmar-Raible, K. (2016). The genomic basis of circadian and circalunar timing adaptations in a midge. *Nature, 540* (7631), 69-73. https://doi.org/10.1038/nature20151

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., & Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Res, 32* (Database issue), 277-280. https://doi.org/10.1093/nar/gkh063

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol, 30* (4), 772-780. https://doi.org/10.1093/molbev/mst010

Kelley, J. L., Peyton, J. T., Fiston-Lavier, A. S., Teets, N. M., Yee, M. C., Johnston, J. S., . . . Denlinger, D. L. (2014). Compact genome of the Antarctic midge is likely an adaptation to an extreme environment.*Nat Commun, 5* , 4611.4611-4611.4618. https://doi.org/10.1038/ncomms5611

Kent, W. J. (2002). BLAT–the BLAST-like alignment tool. *Genome Res, 12* (4), 656-664. https://doi.org/10.1101/gr.229202

Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat Methods, 12* (4), 357-360. https://doi.org/10.1038/nmeth.3317

Kozeretska, I., Serga, S., Kovalenko, P., Gorobchyshyn, V., & Convey, P. (2022). Belgica antarctica (Diptera: Chironomidae): A natural model organism for extreme environments. *Insect Sci, 29* (1), 2-20. https://doi.org/10.1111/1744-7917.12925

Kuranaga, E., Kanuka, H., Igaki, T., Sawamoto, K., Ichijo, H., Okano, H., & Miura, M. (2002). Reaper-mediated inhibition of DIAP1-induced DTRAF1 degradation results in activation of JNK in Drosophila. *Nat Cell Biol, 4* (9), 705-710. https://doi.org/10.1038/ncb842

Kutsenko, A., Svensson, T., Nystedt, B., Lundeberg, J., Bjork, P., Sonnhammer, E., & Giacomello, S. (2014). The Chironomus tentans genome sequence and the organization of the Balbiani ring genes. *BMC Genomics, 15* , 819. https://doi.org/10.1186/1471-2164-15-819

Lacerda, A. C., Gusmao, G. A., & Hamada, N. (2014). Tests of chronic and acute toxicity of crude oil on larvae of Chironomus kiiensis Tokunaga (Diptera: Chironomidae). *Braz J Biol, 74* (3 Suppl 1), 70-77. https://doi.org/10.1590/1519-6984.24012

Li, S., Zhu, S., Jia, Q., Yuan, D., Ren, C., Li, K., Zhan, S. (2018). The genomic and functional landscapes of developmental plasticity in the American cockroach. *Nat Commun, 9* (1), 1008. https://doi.org/10.1038/s41467-018-03281-1

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol, 15* (12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic acids research, 25* (5), 955-964. https://doi.org/10.1093/nar/25.5.955

Lu, K., Song, Y., & Zeng, R. (2021). The role of cytochrome P450-mediated detoxification in insect adaptation to xenobiotics.*Curr Opin Insect Sci, 43* , 103-107. https://doi.org/10.1016/j.cois.2020.11.004

Manni, M., Berkeley, M. R., Seppey, M., Simao, F. A., & Zdobnov, E. M. (2021). BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution* . https://doi.org/10.1093/molbev/msab199

Marcais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers.*Bioinformatics, 27* (6), 764-770. https://doi.org/10.1093/bioinformatics/btr011

Marti'n-Blanco, E., Pastor-Pareja, J. C., & Garci'a-Bellido, A. (2000). JNK and decapentaplegic signaling control adhesiveness and cytoskeleton dynamics during thorax closure in Drosophila. *PNAS, 97* (14), 7888-7893. https://doi.org/10.1073/pnas.97.14.7888

Martin, J. (2017). Chironomus strenzkei Fittkau, 1968 is a junior synonym of C. striatipennis Kieffer, 1910. *CHIRONOMUS Journal of Chironomidae Research, 30* , 19-25. https://doi.org/10.5324/cjcr.v0i30.2190

Martinez-Paz, P. (2018). Response of detoxification system genes on Chironomus riparius aquatic larvae after antibacterial agent triclosan exposures. *Sci Total Environ, 624* , 1-8. https://doi.org/10.1016/j.scitotenv.2017.12.107

Miki, T., Shinohara, T., Chafino, S., Noji, S., & Tomioka, K. (2020). Photoperiod and temperature separately regulate nymphal development through JH and insulin/TOR signaling pathways in an insect. *Proc Natl Acad Sci U S A, 117* (10), 5525-5531. https://doi.org/10.1073/pnas.1922747117

Nawrocki, E. P., & Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics, 29* (22), 2933-2935. https://doi.org/10.1093/bioinformatics/btt509

Neff, R. A., Wang, M., Vatansever, S., Guo, L., Ming, C., Wang, Q., Horgusluoglu-Moloch, E. (2021).

Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Molecular subtyping of Alzheimer's disease using RNA sequencing data reveals novel mechanisms and targets.*SCIENCE ADVANCES, 7* , 5398. https://doi.org/10.1126/sciadv.abb5398

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol, 32* (1), 268-274. https://doi.org/10.1093/molbev/msu300

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol, 33* (3), 290-295. https://doi.org/10.1038/nbt.3122

Pinder, L. c. V. (1986). BIOLOGY OF FRESHWATER CHIRONOMIDAE. *Ann. Rev. Entomol, 31* , 1-23. https://doi.org/10.1146/annurev.en.31.010186.000245

Platzer-Schltz, I., & Welsch, U. (1969). Zur Entstehung und Feinstruktur der peritrophischen Membran der Larven von Chironomus strenzkei Fittkau (Diptera). *Z. Zellforsch, 100* , 594-605. https://doi.org/10.1007/BF00344378

Platzer-Schultz, I. (1970). Zur Atmungsphysiologie verschiedener Entwicklungsstadien von Chironomus strenzkei Fittkau (Diptera). *Z. vergl. Physiologie, 67* , 179-185. https://doi.org/10.1007/BF00298526

Price, A. L., Jones, N. C., & Pevzner, P. A. (2005). De novo identification of repeat families in large genomes.*Bioinformatics, 21 Suppl 1* , 351-358. https://doi.org/10.1093/bioinformatics/bti1018

Rethmeier, N., Seurinck, J., Montagu, M. V., & Cornelissen, M. (1997). Intron-mediated enhancement of transgene expression in maize is a nuclear, gene-dependent process. *The Plant Journal, 12* (4), 895-899. https://doi.org/10.1046/j.1365-313x.1997.12040895.x

Rose, A. B. (2004). The effect of intron location on intron-mediated enhancement of gene expression in Arabidopsis. *The Plant Journal, 40* (5), 744-751. https://doi.org/10.1111/j.1365-313X.2004.02247.x

Rose, A. B., & Beliakoff, J. A. (2000). Intron-Mediated Enhancement of Gene Expression Independent of Unique Intron Sequences and Splicing.*Plant physiology, 122* , 535-542.

Rozewicki, J., Li, S., Amada, K. M., Standley, D. M., & Katoh, K. (2019). MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Res, 47* (W1), 5-10. https://doi.org/10.1093/nar/gkz342

Saffarini, D. A., Trewitt, P. M., Luhm, R. A., & Bergtrom, G. (1991). Differential regulation of insect globin and actin mRNAs during larval development in Chironomus thummi. *Gene, 101* , 215-222. https://doi.org/10.1016/0378-1119(91)90414-7

Santos, C. G., Humann, F. C., & Hartfelder, K. (2019). Juvenile hormone signaling in insect oogenesis. *Curr Opin Insect Sci, 31* , 43-48. https://doi.org/10.1016/j.cois.2018.07.010

Shaikhutdinov, N., & Gusev, O. (2022). Chironomid midges (Diptera) provide insights into genome evolution in extreme environments.*Curr Opin Insect Sci, 49* , 101-107. https://doi.org/10.1016/j.cois.2021.12.009

Smital, T., Luckenbach, T., Sauerborn, R., Hamdoun, A. M., Vega, R. L., & Epel, D. (2004). Emerging contaminants—pesticides, PPCPs, microbial degradation products and natural substances as inhibitors of multixenobiotic defense in aquatic organisms. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 552* (1-2), 101-117. https://doi.org/10.1016/j.mrfmmm.2004.06.006

Smith, R. (2014). Limnology—Inland water ecosystems. *the North American Benthological Society, 21* (2), 346-347.

Sogame, Y., & Kikawada, T. (2017). Current findings on the molecular mechanisms underlying anhydrobiosis in Polypedilum vanderplanki.*Current Opinion in Insect Science, 19* , 16-21. https://doi.org/10.1016/j.cois.2016.10.008

Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res, 34* (Web Server issue), 435-439. https://doi.org/10.1093/nar/gkl200

Strand, K., Knapp, J. E., Bhyravbhatla, B., & Royer, W. E., Jr. (2004). Crystal structure of the hemoglobin dodecamer from Lumbricus erythrocruorin: allosteric core of giant annelid respiratory complexes. *J Mol Biol, 344* (1), 119-134. https://doi.org/10.1016/j.jmb.2004.08.094

Subramanian, B., Gao, S., Lercher, M. J., Hu, S., & Chen, W. H. (2019). Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res, 47* (W1), 270-275. https://doi.org/10.1093/nar/gkz357

Sun, X., Liu, W., Li, R., Zhao, C., Pan, L., & Yan, C. (2021). A chromosome level genome assembly of Propsilocerus akamusi to understand its response to heavy metal exposure. *Mol Ecol Resour, 21* (6), 1996-2012. https://doi.org/10.1111/1755-0998.13377

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc, 7* (3), 562-578. https://doi.org/10.1038/nprot.2012.016

Vicoso, B., & Bachtrog, D. (2015). Numerous transitions of sex chromosomes in Diptera. *PLoS Biol, 13* (4), 1002078. https://doi.org/10.1371/journal.pbio.1002078

Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., & Schatz, M. C. (2017). GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics, 33* (14), 2202-2204. https://doi.org/10.1093/bioinformatics/xxxxx

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., . . . Paterson, A. H. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res, 40* (7), 49. https://doi.org/10.1093/nar/gkr1293

Wu, Z., Li, R., Zhang, Y., & Zhu, L. (2021). Insights into the impacts of bioturbation by multiple benthic organisms on the bioavailability and toxic effects of perfluorooctane sulfonate in sediment. *J Hazard Mater, 420* , 126675. https://doi.org/10.1016/j.jhazmat.2021.126675

Xiao, Y., Xiao, Z., Ma, D., Liu, J., & Li, J. (2019). Genome sequence of the barred knifejaw Oplegnathus fasciatus (Temminck & Schlegel, 1844): the first chromosome-level draft genome in the family Oplegnathidae. *Gigascience, 8* (3). https://doi.org/10.1093/gigascience/giz013

Xu, Z., & Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res, 35* (Web Server issue), 265-268. https://doi.org/10.1093/nar/gkm286

Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *CABIOS APPLICATIONS NOTE, 13* (5), 555-556.

Zeitlinger, J., & Bohmann, D. (1999). Thorax closure in Drosophila: involvement of Fos and the JNK pathway. *Great Britain, 126* , 3947-3956. https://doi.org/10.1007/s004290050284

Zhang, Y., Fu, Y., Wang, Q., Liu, X., Li, Q., & Chen, J. (2020). Transcriptome analysis reveals rapid defence responses in wheat induced by phytotoxic aphid Schizaphis graminum feeding. *BMC Genomics, 21* (1), 339. https://doi.org/10.1186/s12864-020-6743-5

Zhang, Y., Yang, J., Simpson, S. L., Wang, Y., & Zhu, L. (2019). Application of diffusive gradients in thin films (DGT) and simultaneously extracted metals (SEM) for evaluating bioavailability of metal contaminants in the sediments of Taihu Lake, China. *Ecotoxicol Environ Saf, 184* , 109627.109621-109627.109628. https://doi.org/10.1016/j.ecoenv.2019.109627

Zhang, Y., Zhang, G. Q., Zhang, D., Liu, X. D., Xu, X. Y., Sun, W. H., . . . Liu, Z. J. (2021).

Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Chromosome-scale assembly of the Dendrobium chrysotoxum genome enhances the understanding of orchid evolution.*Hortic Res, 8* (1), 183. https://doi.org/10.1038/s41438-021-00621-z

Ethics Statement

This research was approved by the Animal Ethics Committee of Shanghai Ocean University, Shanghai, China. All experiments were performed in strict accordance with the requirements of the Animal Ethics Procedures and Guidelines of the People's Republic of China.

Author Contributions

Bingxin Guo wrote the manuscript; Jinxian Chen performed nanopore sequencing, and transcript sequencing, and functional gene analysis; Liang Lu assembled the genome, and bioinformatic analyses; Li Wang carried out the genome annotation; Chenhong Li performed Illunima sequencing and estimated the genome size; Wenbing Liu assessed the assembly quality; Ruilei Zhang and Liqing Wang designed the projects of this study, collected the samples, and determined *Chironomus striatipennis* . All authors read, edited, and approved the final manuscript.

Table Legends

Table 1. Genome statistics and comparisons among Chironomid species whose genome has been sequenced

Figure Legends

Figure 1. Photographs of *Chironomus striatipennis* larva

Figure 2. Hi-C contact map of the*Chironomus striatipennis* **genome assembly. U, unplaced scaffolds.**

Figure 3. Circular diagram depicting the characteristics of the*Chironomus striatipennis* genome. From the outer to the inner circle: A Ideogram of the 4 *C. striatipennis* chromosomes at the Mb scale. B Guanine-cytosine (GC) content. C Repeat density. D Gene density. E Collinearity block in the genome. Connecting lines at the centre of the diagram highlight the homologous relationships of chromosomes

Figure 4. Orthology and genome evolution of *Chironomus stritipennis* compared with those of 10 other species. The 10 species used for comparison were *C. riparius, C. tentans, C. marinus, P. vanderplanki, P. akamusi, A. gambiae, A. sinensis, C. quinquefasciatus, D. melanogaster and M. domestica.* IQ-TREE was used to construct the unrooted maximum-likelihood phylogenetic tree for the 11 species based on genomic data obtained using LG + F + R5 model for 1000 bootstrap replicates.

Figure 5. Analysis of *Chironomus striatipennis* Globin genes. a. Phylogenetic relationships of *C. striatipennis, C. riparius, C. tentans, C. marinus, P. akamusi and P. vanderplanki* Globin proteins. IQ-TREE was used to build the unrooted maximum-likelihood phylogenetic tree with the VT+R4 Best-fit model for 1000 bootstrap replicates. 10, *C. striatipennis* ; 0, *C. riparius* ; 5,*C. tentans* ; 6, *C. marinus* ; 1, *P. akamusi* ; 7,*P. vanderplanki* . b. Expression patterns of Globin in larva, pupa and adult of *C. striatipennis* .

Figure 6. Gene families involved in CYP450s in *Chironomus striatipennis*. **a** Comparison of gene numbers for Mito, CYP2, CYP3, CYP4 proteins in *C. striatipennis, C. riparius, C. tentans, C. marinus, P. vanderplanki* and *P. akamusi* . Gene numbers are provided above each bar. **b** Phylogeny tree of CYP450s genes from C. *striatipennis, C. riparius, C. tentans, C. marinus, P. akamusi* and*P. vanderplanki* Globin proteins. IQ-TREE was used to build the unrooted maximum-likelihood phylogenetic tree with the LG+R6 Best-fit model for 1000 bootstrap replicates. 10, *C. striatipennis* ; 0,*C. riparius* ; 5, *C. tentans* ; 6, *C. marinus* 1,*P. akamusi* ; 7, *P. vanderplanki.*

Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Figure 7. Gene expression patterns among the interactions between hemoglobin, JH III and Ecdysone biosynthesis pathways, and regulatory and metamorphosed mechanisms in *Chironomus striatipennis*. (Frankenberg et al., 2001; Saffarini et al., 1991; Belles X et al., 2005; Y Demay et al., 2014; Kuranaga et al., 2002)

Figure 8. Genome and transcriptome analyses JHAMT of the gene family in Chironomus striatipennis. Comparison of all transcripts and JHAMT transcripts under different stages (larva, pupa, adult); All transcripts on the left and JHAMT transcripts on the right.

Supplemental information

Supplementary Figure 1. Genome size and heterozygosity estimation using 19 *K-mer*distribution.

Supplementary Figure 2. BUSCO assess the completeness of the genome draft of *C. striatipennis* genome after three polish. (base on the diptera_odb10 databases)

Supplementary Figure 3. BUSCO assess the completeness of the chromosome level genome of *C. striatipennis* genome after annotation. (base on the insecta_odb10, diptera_odb10 and metazoan_odb10 databases)

Supplementary Figure 4. Phylogenetic tree constructed based on the filtered single-copy gene family; each branch length represents the neutral evolution rate.

Note:In this paper, due to the missing of transcripts and gff files of*Belgica antarctica, Chironomus tepperi* and *Polypedilum pembai* , there is no phylogenetic studies.

Supplementary Figure 5. Phylogeny of expanded and contracted gene families. Numbers in the tree illustrate the sets of gene-family expansions (+, red) and contractions (–, blue) found for *C. striatipennis*.

Supplementary Figure 6. Gene Ontology (GO) enrichment results of significantly expanded gene families in *C. striatipennis*.

Supplementary Figure 7. Gene Ontology (GO) enrichment results of significantly shrinked gene families in *C. striatipennis*.

Supplementary Figure 8. Phylogeny tree of GST genes from*C. striatipennis, C. riparius, C. tentans, C. marinus, P. akamusi*and *P. vanderplanki* Globin proteins. IQ-TREE was used to build the unrooted maximum-likelihood phylogenetic tree with the VT+F+G4 Best-fit model for 1000 bootstrap replicates. 10, *C. striatipennis*; 0, *C. riparius*; 5, *C. tentans*; 6, *C. marinus*; 1, *P. akamusi*; 7, *P. vanderplanki*.

Supplementary Figure 9. Phylogeny tree of caspase7 genes from*C. striatipennis, C. riparius, C. tentans, C. marinus, P. akamusi*and *P. vanderplanki* Globin proteins. IQ-TREE was used to build the unrooted maximum-likelihood phylogenetic tree with the VT+F+R2 Best-fit model for 1000 bootstrap replicates. 10, *C. striatipennis*; 0, *C. riparius*; 5, *C. tentans*; 6, *C. marinus*; 1, *P. akamusi*; 7, *P. vanderplanki*.

Supplementary Figure 10. Phylogeny tree of JHAMT genes from*C. striatipennis, C. riparius, C. tentans, C. marinus, P. akamusi*and *P. vanderplanki* JHAMT proteins. IQ-TREE was used to build the unrooted maximum-likelihood phylogenetic tree with the VT+F+R3 Best-fit model for 1000 bootstrap replicates. 10, *C. striatipennis*; 0,*C. riparius*; 5, *C. tentans*; 6, *C. marinus*; 1,*P. akamusi*; 7, *P. vanderplanki*.

Supplementary Figure 11. Phylogeny tree of chitinase genes from*C. striatipennis, C. riparius, C. tentans, C. marinus, P. akamusi*and *P. vanderplanki* chitinase proteins. IQ-TREE was used to build the unrooted maximum-likelihood phylogenetic tree with the VT+F+G4 Best-fit

Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

model for 1000 bootstrap replicates. 10, *C. striatipennis*; 0, *C. riparius*; 5, *C. tentans*; 6, *C. marinus*; 1, *P. akamusi*; 7, *P. vanderplanki*.

Supplementary Table 1. Summary of DNA sequencing data.

Supplementary Table 2. The pairwise genetic distances (%) of*C. striatipennis* Kieffer.

Supplementary Table 3. Summary of nextdenovo assembly results of*C. striatipennis*.

Supplementary Table 4. BUSCO assessment of the *C. riparius* and *C. tentans*genomes' assembly.

Supplementary Table 5. Chromosome length by Hi-C assembly.

Supplementary Table 6. Summary of orthologous gene families in 11 sequenced diptera species.

Supplementary Table 7. BUSCO assessment of the *C. riparius* and *C. tentans* genomes' annotation.

Supplementary Table 8. Genome assembly and annotation summary for *C. striatipennis* and 7 other Chironomidae species.

Supplementary Table 9. Statistics of non-coding RNA in the*C. striatipennis* genome.

Supplementary Table 10. Statistical results of repeat sequences.

Supplementary Table 11. Statistics of repeat sequences in*C. striatipennis*.

Supplementary Table 12. Statistical results of functional annotation.

Supplementary Table 13. Statistical results of whole gene collinearity analysis.

Supplementary Table 14. Summary of RNA sequencing data.

Supplementary Table 15. The expression of genes in TCA pathway.

Supplementary Table 16. The expression of Globin genes in*C. striatipennis*.

Supplementary Table 17. The expression of genes in porphyrin biosynthesis pathway and regulatory mechanisms.

Supplementary Table 18. The expression of genes in hormone biosynthesis pathway and regulatory mechanisms.

Supplementary Table 19. The expression of genes in metamorphosed mechanisms.

Supplementary Table 20. The expression of chitinase genes in*C. striatipennis*.

Table 1. Genome statistics and comparisons among chironomid species whose genome has been sequenced

Species	Total length (Mb)	Total length (Mb)	Total length (Mb)	Gene number	Contig N50 (bp)	Scaffold
C. striatipennis	*C. striatipennis*	181.84	13429	13429	5182970	545130
C. riparius	*C. riparius*	172.34	13449	13449	284852	539778
C. tentans	*C. tentans*	213.46	15120	15120	7697	57274
C. marinus	*C. marinus*	85.49	22767	22767	154800	187115
P. vanderplanki	*P. vanderplanki*	118.97	17852	17852	219078	352090
P. akamusi	*P. akamusi*	85.84	11942	11942	6207813	261102

This study

Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Figure 1. Photographs of *Chironomus striatipennis* larva(Note: Due to the epidemic, it is temporarily impossible to take photos at school, and the specific photos can be changed later)

Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

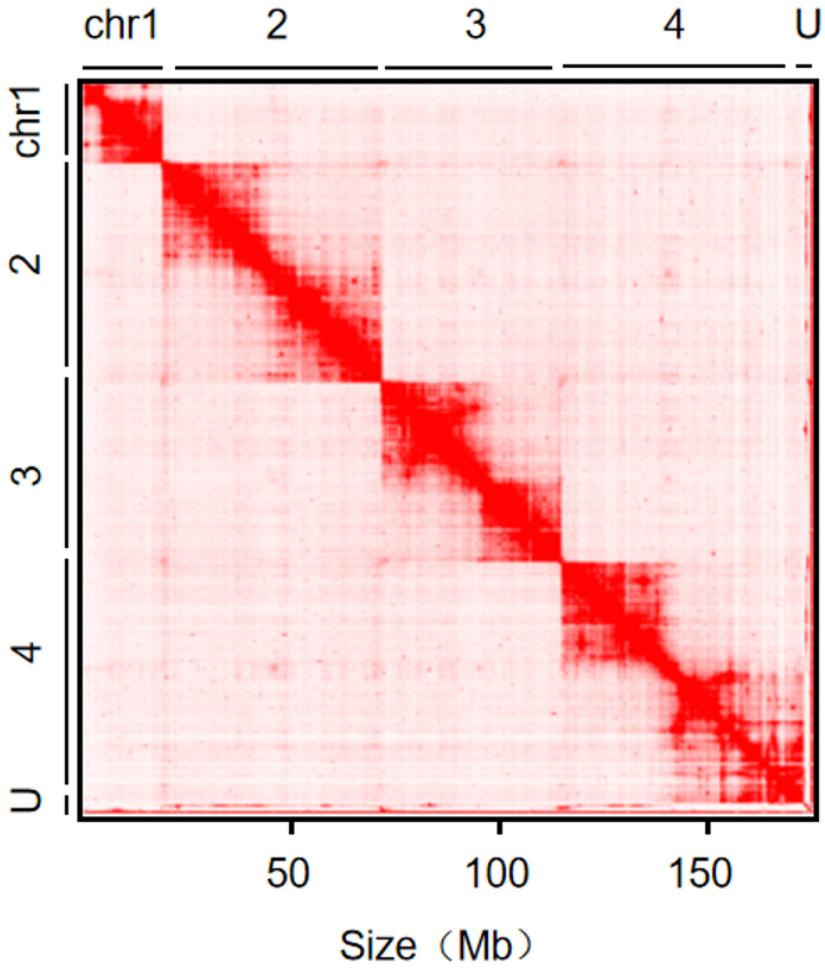


Figure 2. Hi-C contact map of the *Chironomus striatipennis* genome assembly. U, unplaced scaffolds.

Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

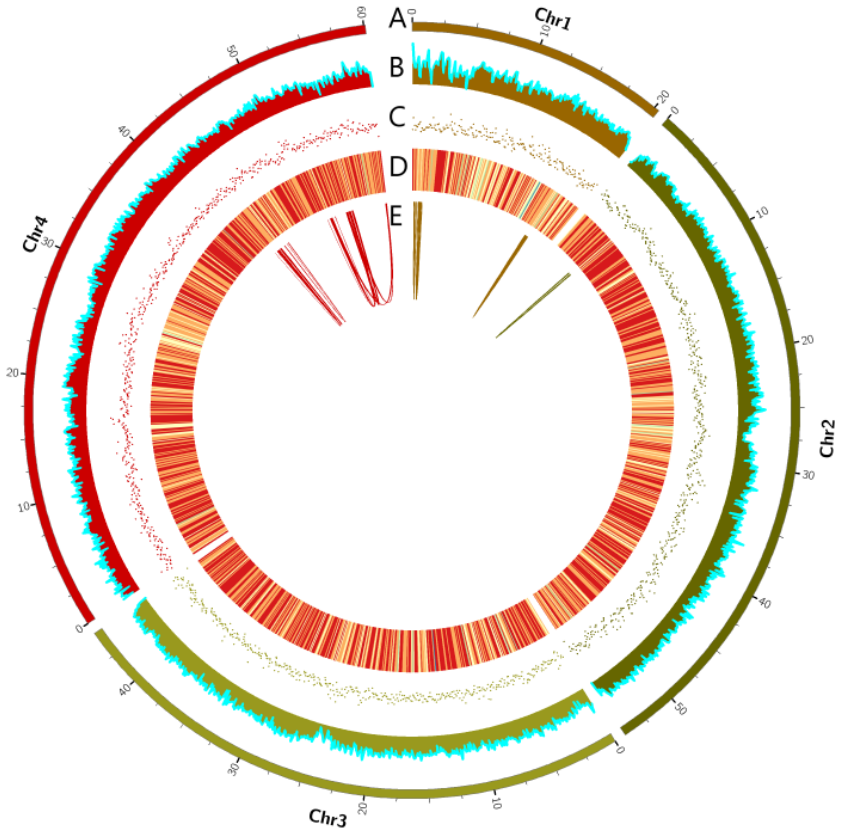
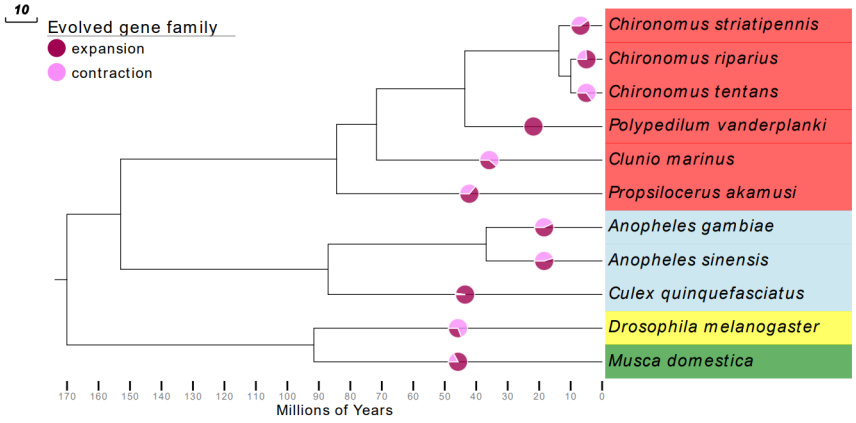
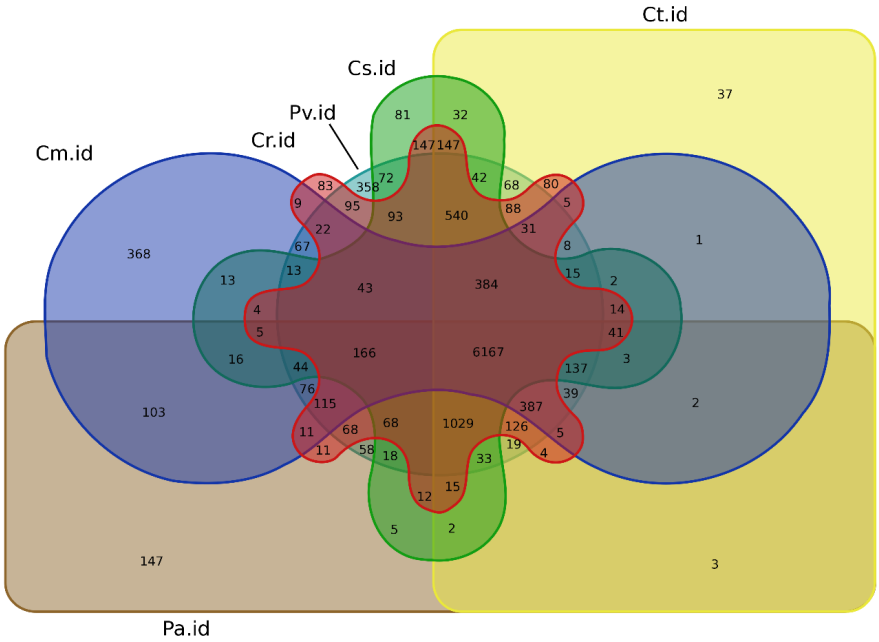


Figure 3. Circular diagram depicting the characteristics of the *Chironomus striatipennis* genome. From the outer to the inner circle: A Ideogram of the 4 *C. striatipennis*chromosomes at the Mb scale. B Guanine-cytosine (GC) content. C Repeat density. D Gene density. E Collinearity block in the genome. Connecting lines at the centre of the diagram highlight the homologous relationships of chromosomes



a

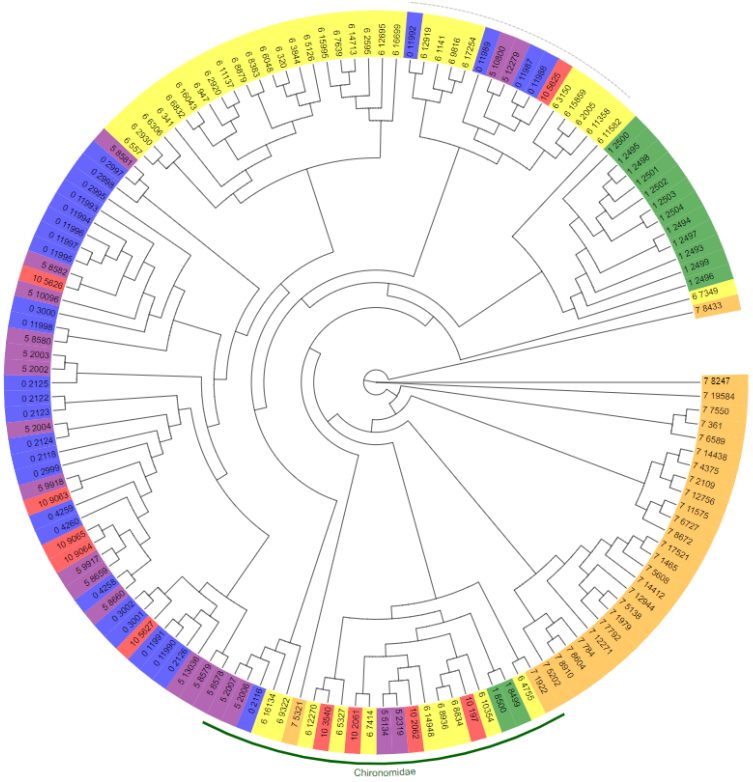


b

Figure 4. Phylogenomic analysis revealed a wide range of genetic changes. a Orthology and genome evolution of *Chironomus stritipennis* compared with those of 10 other species. The 10 species used for comparison were *C. riparius, C. tentans, C. marinus, P. vanderplanki, P. akamusi, A. gambiae, A. sinensis, C. quinquefasciatus, D. melanogaster and M. domestica* . IQ-TREE was used to construct the unrooted maximum-likelihood phylogenetic tree for the 11 species based on genomic data obtained using LG + F + R5 model for 1000 bootstrap replicates. **b** Venn analysis of Orthogroups of the six Chironomid midges. Cs.id, *C. striatipennis* ; Cr.id, *C. riparius* ; Ct.id, *C. tentans* ; Cm.id, *C. marinus* Pa.id,*P. akamusi* ; Pv.id, *P. vanderplanki* .

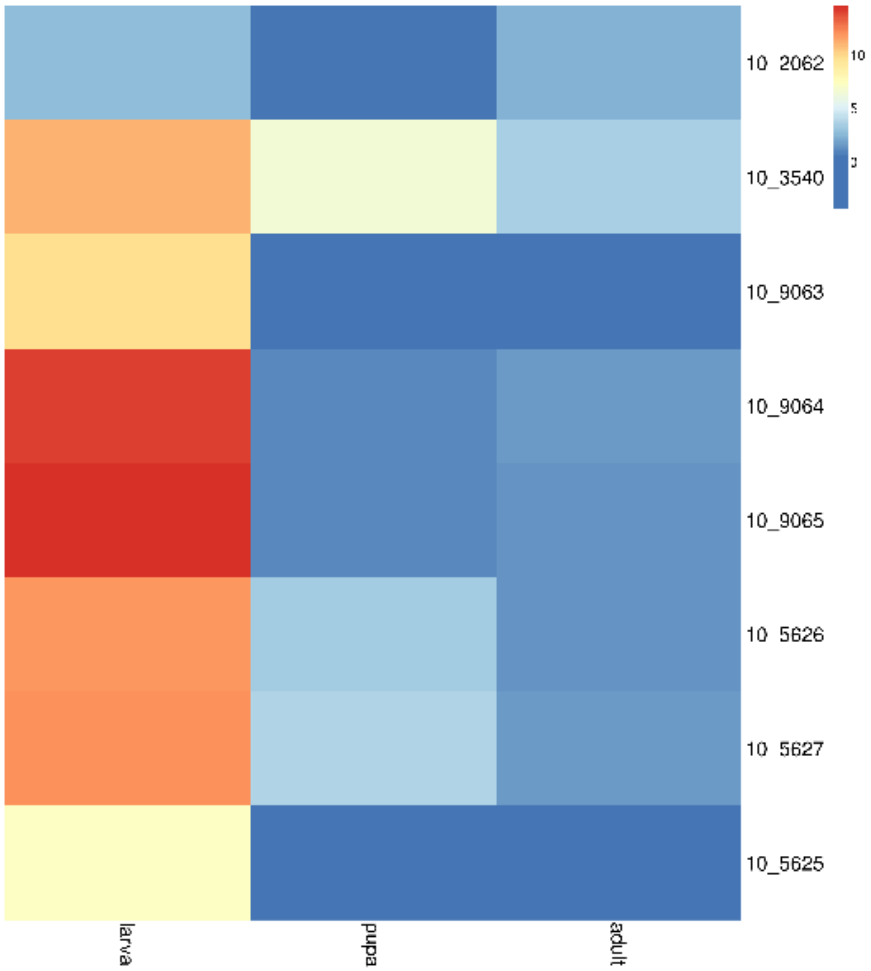
Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709098.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.



a

Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

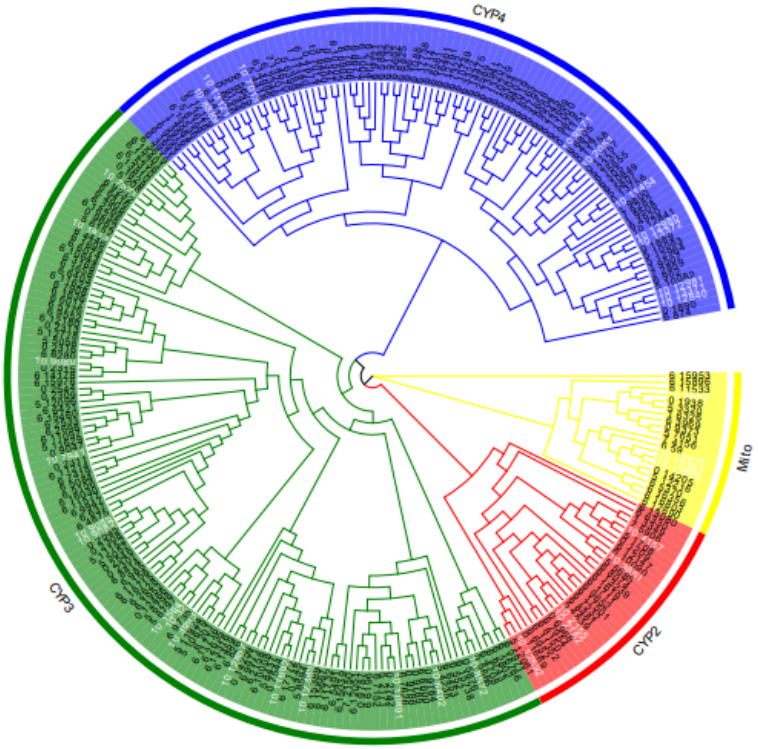


b

Figure 5. Analysis of *Chironomus striatipennis* Globin genes. a. Phylogenetic relationships of *C. striatipennis, C. riparius, C. tentans, C. marinus, P. akamusi and P. vanderplanki* Globin proteins. IQ-TREE was used to build the unrooted maximum-likelihood phylogenetic tree with the VT+R4 Best-fit model for 1000 bootstrap replicates. 10, *C. striatipennis* ; 0, *C. riparius* ; 5,*C. tentans* ; 6, *C. marinus* ; 1, *P. akamusi* ; 7,*P. vanderplanki* . b. Expression patterns of Globin in larva, pupa and adult of *C. striatipennis* .

a

Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.



b

Figure 6. Gene families involved in CYP450s in *Chironomus striatipennis*. **a** Comparison of gene numbers for Mito, CYP2, CYP3, CYP4 proteins in *C. striatipennis, C. riparius, C. tentans, C. marinus, P. vanderplanki* and *P. akamusi* . Gene numbers are provided above each bar. **b** Phylogeny tree of CYP450s genes from C. *striatipennis, C. riparius, C. tentans, C. marinus, P. akamusi* and*P. vanderplanki* CYP450s proteins. IQ-TREE was used to build the unrooted maximum-likelihood phylogenetic tree with the LG+R6 Best-fit model for 1000 bootstrap replicates. 10, *C. striatipennis* ; 0,*C. riparius* ; 5, *C. tentans* ; 6, *C. marinus* ; 1,*P. akamusi* ; 7, *P. vanderplanki.*

Posted on Authorea 6 Jul 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.165709988.84463352/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

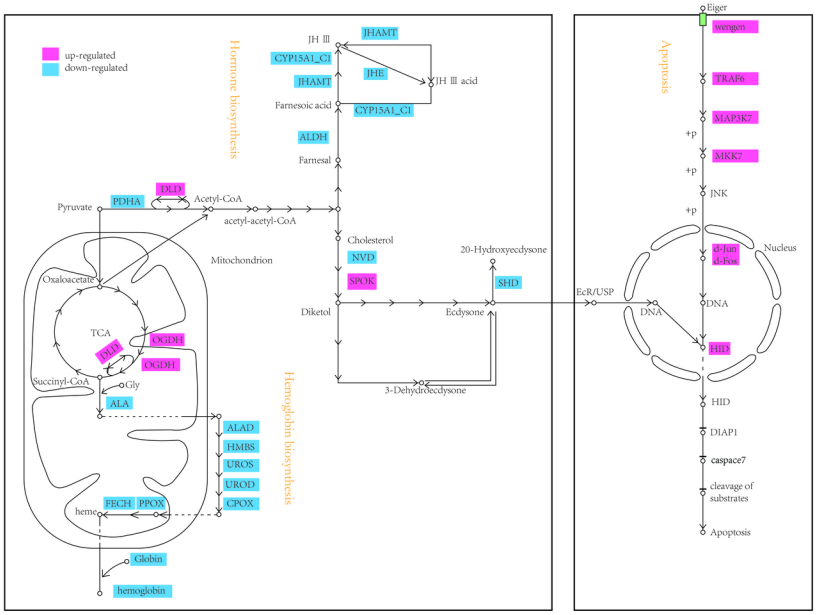


Figure Figure 7. Gene expression patterns among the interactions between hemoglobin, JH III and Ecdysone biosynthesis pathways, and regulatory andmetamorphosed mechanisms in *Chironomus striatipennis*.(Frankenberg et al., 2001; Saffarini et al., 1991; Belles X et al., 2005; Demay, Y., et al., 2014; Kuranaga et al., 2002) Increased gene expression during the developmental stages of midges shown by red backgrounds, while decreased gene expression is shown by blue backgrounds. The pathways of hemoglobin synthesis, hormone synthesis and apoptosis were marked.

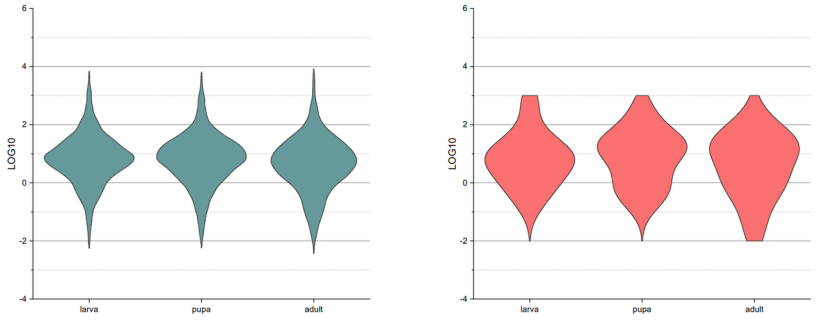


Figure 8. Genome and transcriptome analyses JHAMT of the gene family in Chironomus striatipennis. Comparison of all transcripts and JHAMT transcripts under different stages (larva, pupa, adult); All transcripts on the left and JHAMT transcripts on the right.