A chromosome-level genome supplies an insight into the distinctive diapause characteristics in Caligula japonica

Xu Chen[1], Yong-Ming Chen[1], Su Wang[2], Kang He[3], Mengyao Chen[3], and Lian-Sheng Zang[1]

[1]Guizhou University
[2]Beijing Academy of Agricultural and Forestry Sciences
[3]Zhejiang University

November 4, 2022

Abstract

Caligula japonica is a forestry pest due to its damage to multiple trees. Recently, it served as a potential natural medical mesh biomaterial in the medical industry. However, studies on karyotype evolution and functional genomics of C. japonica are limited by the absent of genomic resource and its several months of diapause. Here, we conducted high-throughput sequencing of its genome and firstly obtained the chromosome-level genome. we successfully assembled a high-quality genome of 584,506,556bp with Contig N50 of 12 Mb and 31 chromosomes. About 342 Mb repeat sequences were identified, accounting for 58.53% of C. japonica genome. Genome annotation by de novo gene prediction and homologous gene search yields 24791 protein-coding genes. We also applied manual annotation of diapause genes in C. japonica genome. The genome abstained in this research will not only support resource for future study on diapause mechanism of C. japonica but also help to progress comparative genomic analyses in Lepidoptera.

1 INTRODUCTION

The Japanese giant silkworm, *Caligula japonica* (Moore 1872) (Lepidoptera: Bombycoidea), is generally considered as a forestry pest due to its damage to many trees, particular for walnut tree (Chen*et al* ., 2021a). It is widely distributed in Asian, and its damage caused huge economic loss every year in Japan, North Korea, Russia, and especially as to China. However, recent research indicated that the cocoon of pupa (Figure 1) displays nontoxicity, biocompatibility, suitable mechanical properties, and porosity while showing no adverse effect in animal trials and even appears to enhance cell proliferation, so it could be used as natural medical mesh in the medical industry (Chen *et al.* , 2022). While the genetical mechanism of this compatibility are completely unexplored. There are two types of diapause for *C. japonica* . It overwinters with egg diapause. The egg period lasts until the middle of May of the next year and pupates from the middle of July. The pupal period is about 40 days, which shows summer diapause (Chen *et al.* , 2021b). Diapause of insects is common, but it is rare for a species to have two diapause modes. At molecular level, the regulation mechanism related to diapause *C. japonica* have remained unknown. To more in-depth research on regulation of diapause and exert its potential medical value of *C. japonica* , a high-quality genome is an essential tool.

As a miniature animal group on the earth, insects have evolved a variety of strategies to resist unfit environment to survive, among that diapause is a very effective strategy. The regulation of insect diapause involves many factors, such as the induction of abiotic factors, photoperiod, temperature, and humidity. Both long and short light can induce insects to enter diapause, For instance, *Mamestra brassicae* will enter summer diapause when the illumination is $>$ 12h (Goto and Hukushima, 1995); *Chilo suppressalis*will enter winter

diapause when the illumination is less than 12h (Xiao*et al.* , 2010). While the regulation of hormones and some foundational proteins are always inseparable as biological factor (Denlinger, 2002). Studies have shown that Ecdysone, juvenile hormone, prothoracic tropic hormone (PTTH) and some rhythm genes,including diapause hormone (*DH*), period (*per*) , timeless (*tim*) , Cryptochrome 1(*CRY 1*), Cryptochrome 2(*CRY 2*), can affect the diapause of insects (Denlinger, 1985). Nevertheless, it is poorly understood how rhythm genes regulate diapause of *C. japonica* .

The high-quality genome can help to explore the new function of key genes in insect physiological regulation and searching for new metabolic pathways. Here, to better study the diapause mechanism and explore the genetic of *C. japonica* , we conducted high-throughput sequencing of its genome and obtained high-quality chromosome level genome. The rhythm genes closely related to diapause were annotated manually. We provide a helpful genome sequence resources for in-depth investigations for insect diapause and forward research on *C. japonica* .

2 MATERIALS AND METHODS

2.1 Insects collection

The *C. japonica* were fed on walnut tree growing naturally. We collected the pupae from walnut trees on mountain from Longnan, Gansu province, China in July 2019. Then pupae were placed in Key Laboratory of Green Pesticide and Agricultural Bioengineering of Ministry of Education, Guizhou University, under 27±2, relative humidity of 75+-5%, and photoperiod 14 L: 10D for 40 days. After emerging, 1 male was collected and stored at -80 for sequencing.

2.2 Genome size estimation by flow cytometry

Flow cytometry was used to estimate the genome size of *C. japonica* according to the standard procedure (He *et al.* , 2016). One head of male adult was homogenized completely with 500 μL ice-cold Galbraith's buffer (PH=7). The homogenate was centrifuged at 5,000 rpm. at 25 for 5 min and suspended with 400 μL phosphate buffer (PH=7.4). To remove the RNA, RNaseA was add into tube at 25 for 10 min (final concentration of 20 ug/mL). Finally, samples were stained with 50 μg/mL propidium iodide stock solution in darkness at 4 for 10 min. Samples were analyzed by FACSCalibur platform (BD Biosciences) and FACScomp v4.0 under 488-nm wavelength. FlowJo v7.6.1 was used to obtain the nuclei peaks. *Drosophila melanogaster* was analyzed as a control following the above same parameter. The outputs were used to estimate the genome size.

2.3 Genome sequencing and assembly

Illumine sequencing was performed to evaluate genome size, heterozygosity and rate of duplication and polish de novo assembly. A paired-end library (insert size: 350 bp) was constructed on Illumina NovaSeq platform. The raw data generated were filtered by FASTAQC. After filtering, we yielded a total of 112.81 Gb clean data with 176× sequence coverage.

High-quality genome DNA was fragmented to construct SMRT bell library with PCR-free. After the library size was tested to be qualified by Qubit 3.0 and Agilent 2100, it was sequenced on SMRT cell by PacBio Sequel II sequencing platform (Pacific Biosciences) with x186.17 Mean Depth. we obtained a total of 169.37 Gb clean data after filtering and 7,960,820 subreads (mean subreads length: 21,275.65 bp, subreads length N50: 31,540 bp). Row data generated from PacBio sequencing were corrected by CANU. In the assembly phase, reads were assembled into contig and output consensus sequences by WTDBG v2 with default parameters. PBMM2 (MINIMAP2) was used to map original data to the reference genome, and ARROW (RACON) for polishing. The previously polished FASTA sequence was indexed with BWA index, and the corrected genome was used as the reference genome. Then, the Illumina sequencing FASTQ data were compared with BWA MEM to perform Pilon error correction for secondary polishing. To remove the redundancy of the genome after preliminary assembly and error correction, PURGE_HAPLOTIGS software was used to identify and remove the redundant heterozygous contigs according to the depth distribution of reads and sequence

similarity. The quality of genome sequence was evaluated by BUSCO v4 with default parameters(Manni *et al.* , 2021).

2.4 Hi-C assisted assembly

To obtain chromosome-level genome, we used (High-throughput/resolution chromosome conformation capture) Hi-C technology to assist assembly. The eggs were treated by paraformaldehyde to fixed DNA conformation for 10 mins and stopped crosslinking by 2.5 M glycine for 20 mins. Crosslink DNAs were cut with restriction enzyme and produced fill ends with biotin, which were used to build library and subsequent sequencing via Illumina platform. High quality clean data 58.615 Gb (read length: 150 bp) was generated after sequencing and filtering, then used to preliminary assembly by applying 3D-DNA pipeline using default parameters. Using Juicer to build the chromosome interaction map and using Juicebox to visually correct it for finding mistakes of contig sequence, direction or assembly error inside contig.

2.5 Annotation of repeats

The repeat sequences of *C. japonica* were marked by the combining of RepeatModeler v2.0.2 and RepeatMasker v4.1.2 (http://www.repeatmasker.org/). Firstly, a de novo repeat database was built by RepeatModeler. Then using RepeatMasker to perform the repeat sequence annotation based on the Repbase (*https://www.girinst.org/server/RepBase/index.php*). The random repeats were detected by TRF (Tandem Repeats Finder) v4.09 (Benson, 1999). And the sequence search engine was RMBlast v2.11.0 (http://www.repeatmasker.org/RMBlast.html) for sequence alignment.

2.6 Gene prediction and annotation

de novo prediction and homologous gene search were used to protein-coding gene annotation in the *C. japonica* genome. Repeat-masked genome were used to subsequent analysis according to the EVidenceModeler (EVM) v1.1.1 genome annotation pipeline (Haas *et al.* , 2008). First, we used BRAKER v2 (https://github.com/Gaius-Augustus/BRAKER) to perform de novo gene prediction. Second, the protein sequences of Lepidoptera insect were downloaded from NCBI RefSeq as templates for homologous-based predictions by GenomeThreader v 1.7.3 (Gremme *et al.* , 2005). Finally, EVidenceModeler was used to integrate the above two evidence with different weights and obtained the GFF3 format files. The number of genes will be annotated finally.

2.7 Comparative genomics analysis

OrthoFinder v2.5.1 (Emms and Kelly, 2015, 2019) was used to analysis the orthologous and paralogous genes of 10 insect genomes including *D. melanogaster* (assembly accession: GCF_000001215.4)*, Plutella xylostella* (assembly accession: GCA_019096205.1), *Danaus plexippus* (assembly accession: GCF_009731565.1), *Antheraea yamamai* (Kim. Et al., 2018), *B. mori* (assembly accession: GCF_014905235.1), *A. pernyi* (assembly accession: GCA_015888305.1), *C. japonica* (in this study), *Samia ricini* (assembly accession: GCA_014132275.1), *Spodoptera frugiperda* (assembly accession: GCF_011064685.1), *H. armigera*(assembly accession: GCF_002156985.1). And *D. melanogaster* was selected as outgroup.

2.8 Phylogenetic analysis

1167 single-copy Orthologues shared by 10 insects were used to phylogenetic analysis. Supergenes formed by multiple alignment of single copy gene family were constructed for tree construction. The phylogenetic tree was constructed by maximum likelihood (ML) using IQ-tree with the best model (JTT+F+R5) and 1000 rapid bootstrap replicates. Another method to construct phylogenetic tree was used Astral-III to merge all gene trees obtained by OrthoFinder into one species tree. Two trees from above methods must be coincident. Divergence time was estimated by MCMCtree program in the PAML package v4.8 based on the multiple sequence alignment protein sequences. The calibration time points of *P. xylostella* (286 MYA), *D. plexippus* (179 MYA), *S. frugiperd* (140 MYA), *B. mori* (108MYA) and *H. armigera* (56 MYA) were obtained from TimeTree (*http://timetree.org/*). The gene family contraction and expansion were analyzed by CAFE v4.2

based on the results from orthofinder and phylogenetic tree with divergence time (Han*et al.* , 2013). Finally, Evolview (*http://evolgenius.info/#/*) was used to beautify the phylogenetic tree.

2.9 Manual annotation of diapause genes

To annotate the genes involved to diapause, we used the protein sequence of DH-PBAN, PTTH, , CRY 1, CRY 2, PER and TIM from Uniport (*https://www.uniprot.org/help/uniprotkb*) and NCBI (https://www.ncbi.nlm.nih.gov/) as queries to blast against the protein sequence of *C. japonica* and other 9 insects by BLASTP v2.8.1 (evalue: 1e-5). The prediction sequences from BLASTP were checked by Hmmsearch v3.3.1 based on domains loaded from Pfam (*http://pfam.xfam.org*) (DH: PF05874, CRY 1&, CRY 2: PF03441, PER: PF12114, TIM: PF04821) (Potter *et al.* , 2018).

3 RESULTS

3.1 Genome assembly

The genome size of *C. japonica* was estimated to be 504 Mb by flow cytometry, which is consistent with that estimated by K-mer analysis (K=35) using Illumina short reads (Figure 2a). The heterozygosity is 0.47%.

With 169.37 Gb PacBio long reads, we successfully assembled a high-quality genome of 584,506,556bp, with Contig N50 of 12 Mb. A total of 771 contigs were obtained, with the longest contig as 24,253,087 bp. BUSCO analysis with insecta_ODB10 showed that the gene space is 95.6 % of complete genes, suggesting the assembled genome is of high quality and suitable for further analysis. (Table 1). A total of 97.15% assembled genome sequences were anchored to 31 chromosomes by Hi-C scaffolding, with Scaffold N50 of 20,239,873bp (Figure 2b).

3.2 Genome annotation

In total, 342 Mb repeat sequences were identified, accounting for 58.53% of *C. japonica* genome, which was similar as the close relatives *A. pernyi* (60.74%), but higher than *A. yamamai* (37.33%) and *S. ricini* (34.3%) (Table S1). The Unclassified, long interspersed nuclear elements (LINEs) and DNA transposons are the most abundant TEs. Genome annotation by *de novo* gene prediction and homologous gene search yields 24791 protein-coding genes, much more than the number in *A. pernyi*(20814), *A. yamamai* (14638) and *S. ricini* (20366) (Table 2).

3.3 Phylogenetic analysis

To infer the evolutionary status and trace the phylogenetic placement of*C. japonica* , we did phylogenetic analysis for 10 insects (*D. melanogaster* , *P. xylostella* , *D. plexippus* ,*A. yamamai* , *B. mori* , *A. pernyi* , *C. japonica* ,*S. ricini* , *S. frugiperda* , *H. armigera*). 177071 genes were assigned by OrthoFinder, and 158040 genes were in orthogroups (89.3%) of 10 insects compared in our analyses. The total number of orthogroups is 17425, and there are 4382 orthogroups (25%) shared by all ten tested insects. Next, we used 1167 single-copy genes identified by OrthoFinder to the construct a phylogenetic tree, showing that *C. japonica* is diverged from Antheraea about 42.7 million years ago. There are 1812 genes expanding and 1665 genes contracting.

3.4 Manual Annotation of circadian genes of *C. japonica*

To explore the potential mechanism of diapause of *C. japonica* , we manually annotated two diapause related genes (*DH-PBAN* and*PTTH*) and four circadian rhythm genes (*cry1* , *cry2* ,*per* and *tim*) in *C. japonica* and other nine insects (Table 3). *cry1* has two copies in *S. frugiperda* , but only one copy in other insects. We did not annotate *DH-PBAN* and*PTTH* in *D. plexippus* and *P. xylostella* , while*PTTH* is present in almost all Bombycoidea insects, except for*A. yamamai* . It is worth noting that *S. ricini* and *C. japonica* do not have *per* and have the same circadian genes

4 DISCUSSIONS

C. japonica is a kind of forestry pest and has been evaluated for its potential medical value (Chen *et al.* , 2022). It is meaning to turn this pest, which threatens the development of economic forestry, into a useful resource through in-depth study of its genetic and physiological characteristics, which may help us to promote efficient management solutions and develop a massive rearing technique for obtaining abundant population of *C. japonica* as economic resources. In this study, we obtained high-quality genome at chromosome level. With the combination of Illumina NovaSeq, PacBio Sequel II and Hi-C technology, we obtained that *C. japonica* has 31 chromosomes, and the genome size is 584.5 MB.

To annotate the genes involved to diapause, we manually annotated the existence of six potential diapause related genes in *C. japonica*(*DH-PBAN* ,*PTTH, cry1* , *cry2* , *per* and *tim*). The absence of *per* in *C. japonica* attracted our attention. It is not a unique event, although the DNA sequence of clock genes have been well-conserved. It was also not found of *per* in *Cydia pomonella* and *Anoplophora glabripennis* (Cao *et al.* , 2021). The absence of *cry2* in *Drosophila* and the absence of cry1 in *Tribolium castaneum* and *Apis mellifera* has been reported (Yuan *et al.* , 2007) (Sandrelli *et al.* , 2008).*per* is not a necessary gene for these Lepidoptera insects and may have been lost in the process of evolution. *per* and *tim*are transcribed promoted by CLK and CYC in the nucleus and translated into proteins in the cytoplasm. At the same time, the concentration of PER and TIM will also reduce the level of CLK and CYC by negative feedback, thus reducing the transcription of *per* and *tim*(Tomioka and Matsumoto, no date). However, CRY2 is likely more important negative regulator for transcription of CLK and CYC in some insects and can degrade by light. Therefore, it could be existing some complementary factor of PER in function of circadian regulation pathway. Although,*per* gene has occurred mutation in evolution, some non-PER insect can develop normally and adapt rhythmic change regularly. *S. ricini* has the same number of diapause-related genes (in this study) with *C. japonica* and both absence of *per* . However,*A. pernyi* and *A. yamamai* have different related-diapause gene with two formers. Based on their phylogenetic relationship (Figure 4), it may occur some evolutionary events in rhythmic system. In some*per* -null flies, they also can into diapause, that suggests per may not involve in initiating by photoperiodic diapause(Saunders, Henrich and Gilbert, 1989).

Genes related to circadian clock are usually expressed in neurosecretory cells in the brain, and these regions also regulate diapause-related hormones gene expression(Meuti and Denlinger, 2013). In non-diapause mutation files of *Chymomyza costata* , there is a mutation in 5' leader sequence of *tim* gene, which suggested TIM plays an important role in it regulation for diapause (Pavelka, Shimada and Kostal, 2003). When *per* and *cry2* were knocked down in bean bug, *Riptortus pedestris* , it caused the ovarian development of female adults which should have entered diapause in short photoperiod. Similarly, RNA interference (RNAi) was used to interrupt expression of per and cry2 in male adult, which can avoid diapause (Ikeno, Numata and Goto, 2011). Research above indicate that circadian genes can involve in photoperiodic diapause. Hence, for the development of diapause related research of *C. japonica* , it is important to understand the pathway patterns of circadian genes and diapause-related genes in the genome of the species.

Diapause of resources insects and nature enemies is an important but unaddressed issue. It is a key technical barrier restricting the quantity of products and industrial development in the utilization of resource insects. For example, in the silk industry, the diapause of silkworm will affect the quality and yield of silk. Therefore, the diapause removal of silkworm and its related regulation mechanism have been studied very deeply. If the female adult experiences high temperature during the embryonic period, it will produce diapause eggs (Tsuchiya *et al* ., 2020). In biological control, the simple acquisition of alternative hosts is the key to the large-scale propagation of natural enemy insects. Diapause also exists in the industrial production of *A. pernyi* , an important intermediate host of natural enemies such as Trichogrammatidae and Eupelmidae. However, the commercial *A. pernyi* has been cultivated, there are still some limitations in quality and yield. Not only Lepidoptera, but other natural enemy insects will also be refrigerated to facilitate storage and transportation in biological control, which will cause diapause. The common problem is that the natural enemy insects take a long time to remove diapause, which will affect their control effect on pests. The cocoon of pupa of *C. japonica* can used to produce natural material. So, it has great potential to explore as a resource insect for its medical value. It is meaningful to know how to regulate the diapause of *C. japonica.*

The evolution and fluctuation of expression of the gene according to diapause usually caused by environment. However, the critical values of natural conditions for diapause of *C. japonica* , such as photoperiod and temperature, have not been explored clearly. Therefore, we will quantify the impact of these abiotic factors on diapause and to explore the gene foundations according to biological changes. The expression of each gene in the *C. japonica* still needs to be explored clearly to better carry out the follow-up verification experiment.

In summary, we firstly reported a high-quality chromosome scale genome of *C. japonica* and discuss the evolution of diapause related genes. It will provide a valuable database for the future study around resource insect and biomedical materials development.

ACKNOWLEGEMENTS

This research was partially supported by the National Natural Science Foundation of China (32172469), Foundation of Postgraduate of Guizhou Province (YJSKYJJ [2021]041), and Natural Science Research Program of Guizhou University (202202).

CONFLICT OF INTEREST

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

L.-S. Z. and S. W. designed the experiments. C. X. and Y.-M. C. collected and provided the moth samples. X. C., M. C. and K. H. analyzed genome data. X. C. wrote the manuscript. S. W. and L. Z. revised the manuscript. All authors read and approved the final manuscript.

Data Accessibility

The genome sequence and assembly in this study have been deposited in NCBI under Bioproject ID PRJNA814848. The raw reads produced by genome assembly (Hi-c) were deposited in Sequence Read Archive (SRA) under accession SRR19719011.

REFERENCES

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic acids research* , 27(2), 573–580.

Cao, L. J., Song, W., Yue, L., Guo, S. K., Chen, J. C., Gong, Y. J., Hoffmann A. A., & Wei, S. J. (2021). Chromosome-level genome of the peach fruit moth *Carposina sasakii*(Lepidoptera: Carposinidae) provides a resource for evolutionary studies on moths. *Molecular Ecology Resources* , 21(3), 834–848. https://doi.org/10.1111/1755-0998.13288.

Chen, Y.-M., Pekdemir, S., Bilican, I., Koc-Bilican, B., Cakmak, B., Ali, A., Zang, L.-S., Onses, M., & Kaya, M. (2021a). Production of natural chitin film from pupal shell of moth: Fabrication of plasmonic surfaces for SERS-based sensing applications.*Carbohydrate Polymers* , 262, 117909. https://doi.org/10.1016/j.carbpol.2021.117909.

Chen, Y.-M., Qu, X.-R., Li, T.-H., Iqbal, A., Wang, X.-G., Ren, Z.-Y., Desneux, N., & Zang, L.-S. (2021b). Performances of six eupelmid egg parasitoids from China on Japanese giant silkworm Caligula japonica with different host age regimes. *Journal of Pest Science* , 94, 309–319.

Chen, Y.-M., Zang, L.-S., Koc-Bilican, B., Bilican, I., Holland, C., Cansaran-Duman, D., Karaduman, T., Colak, A., Bayır, Y., Halici, Z., Ozmen, S., Ali, A., Labidi, J., Elbuken, C., & Kaya, M. (2022). Macroporous Surgical Mesh from a Natural Cocoon Composite. *ACS Sustainable Chemistry & Engineering* , 10(18), 5728–5738. https://doi.org/10.1021/acssuschemeng.1c06941.

Denlinger, D. L. (1985). Hormonal control of diapause.*Comprehensive Insect Physiology, Biochemistry and Pharmacology* , 8, 353–412.

Denlinger, D. L. (2002). Regulation of diapause. *Annual Review of Entomology* , 47(1), 93–122.

Emms, D. M., & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* , 16(1), 1–14. https://doi.org/10.1186/s13059-015-0721-2.

Emms, D.M. & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* , 20(1), 1–14. https://doi.org/10.1186/s13059-019-1832-y.

Goto, M., & Hukushima, H. (1995). Factors affecting the induction of summer and winter-diapause and diapause sensitive larval stage of cabbage-armyworm [Brassica oleracea capitata] moth *Mamestra brassicae* (Lepidoptera: Noctuidae). *Journal of the Yamagata Agriculture and Forestry Society (Japan)* , 52, 17-23.

Gremme, G., Brendel, V., Sparks, M. E., & Kurtz, S. (2005). Engineering a software tool for gene structure prediction in higher organisms.*Information and Software Technology* , 47(15), 965–978. https://doi.org/10.1016/j.infsof.2005.09.005.

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., White, O., Buell, R. C., & Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology* , 9(1), R7. https://doi.org/10.1186/gb-2008-9-1-r7.

Han, M. V., Thomas, G. W., Lugo-Martinez, J., & Hahn, M. W. (2013). Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Molecular Biology and Evolution* , 30(8), 1987–1997. https://doi.org/10.1093/molbev/mst100.

He, K., Lin, K., Wang, G., & Li, F. (2016). Genome sizes of nine insect species determined by flow cytometry and k-mer analysis. *Frontiers in Physiology* , 7, 569. https://doi.org/10.3389/fphys.2016.00.

Ikeno, T., Numata, H., & Goto, S. G. (2011). Photoperiodic response requires mammalian-type cryptochrome in the bean bug Riptortus pedestris. *Biochemical and Biophysical Research Communications* , 410(3), 394–397. https://doi.org/10.1016/j.bbrc.2011.05.142.

Manni, M., Berkeley, M. R., Seppey, M., & Zdobnov, E. M. (2021). BUSCO: assessing genomic data quality and beyond. *Current Protocols* , 1(12), e323. https://doi.org/10.1002/cpz1.323.

Mei, Y., Jing, D., Tang, S., Chen, X., Chen, H., Duanmu, H., Cong, Y. Y., Chen, M.Y., Ye, X. H., Zhou, H., He, K., & Li, F. (2022). InsectBase 2.0: a comprehensive gene resource for insects. *Nucleic Acids Research* , 50(D1), D1040–D1045. https://doi.org/10.1093/nar/gkab1090.

Meuti, M. E., & Denlinger, D. L. (2013). Evolutionary links between circadian clocks and photoperiodic diapause in insects.*Integrative and Comparative Biology* , 53(1), 131–143. https://doi.org/10.1093/icb/ict023.

Pavelka, J. A., Shimada, K. I., & Kostal, V. L. (2003). TIMELESS: A link between fly's circadian and photoperiodic clocks? *European Journal of Entomology* , 100(2), 255–265.https://doi.org/10.14411/eje.2003.041.

Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., & Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Research* , 46(W1), W200–W204. https://doi.org/10.1093/nar/gky448.

Sandrelli, F., Costa, R., Kyriacou, C. P., & Rosato, E. (2008). Comparative analysis of circadian clock genes in insects. *Insect Molecular Biology,* 17(5), 447–463. https://doi.org/10.1111/j.1365-2583.2008.00832.x.

Saunders, D. S., Henrich, V. C., & Gilbert, L. I. (1989). Induction of diapause in Drosophila melanogaster: photoperiodic regulation and the impact of arrhythmic clock mutations on time measurement.*Proceedings of the National Academy of Sciences* , 86(10), 3748–3752. https://doi.org/10.1073/pnas.86.10.3748.

Tomioka, K., & Matsumoto, A. (2010). A comparative view of insect circadian clock systems. *Cellular and Molecular Life Sciences* , 67(9), 1397-1406. https://doi.org/10.1007/s00018-009-0232-y.

Tsuchiya, R., Kaneshima, A., Kobayashi, M., Yamazaki, M., Takasu, Y., Sezutsu, H., Tanaka, Y., Mizoguchi, A., & Shiomi, K. (2021). Maternal GABAergic and GnRH/corazonin pathway modulates egg diapause

phenotype of the silkworm *Bombyx mori* . *Proceedings of the National Academy of Sciences* , 118(1), e2020028118. https://doi.org/10.1073/pnas.2020028118.

Xiao, H. J., Mou, F. C., Zhu, X. F., & Xue, F. S. (2010). Diapause induction, maintenance and termination in the rice stem borer*Chilo suppressalis* (Walker). *Journal of Insect Physiology* , 56(11), 1558–1564. https://doi.org/10.1016/j.jinsphys.2010.05.012.

Yuan, Q., Metterville, D., Briscoe, A. D., & Reppert, S. M. (2007). Insect cryptochromes: gene duplication and loss define diverse ways to construct insect circadian clocks. *Molecular Biology and Evolution* , 24(4), 948–955. https://doi.org/10.1093/molbev/msm011.

TABLE 1 Summary of chromosome-level assembly for *Caligula japonica.*

Genome size (bp)	584,537,256
No. of chromosome	31
No. of contig	771
Chromosome-level Contig N50 (bp)	12,646,762
Chromosome-level Scaffold N50(bp)	20,239,873
BUSCO genes (%)	95.6
Heterozygosity (%)	0.47
Repeat (%)	58.53
G+C (%)	35.13
No. of genes	24791

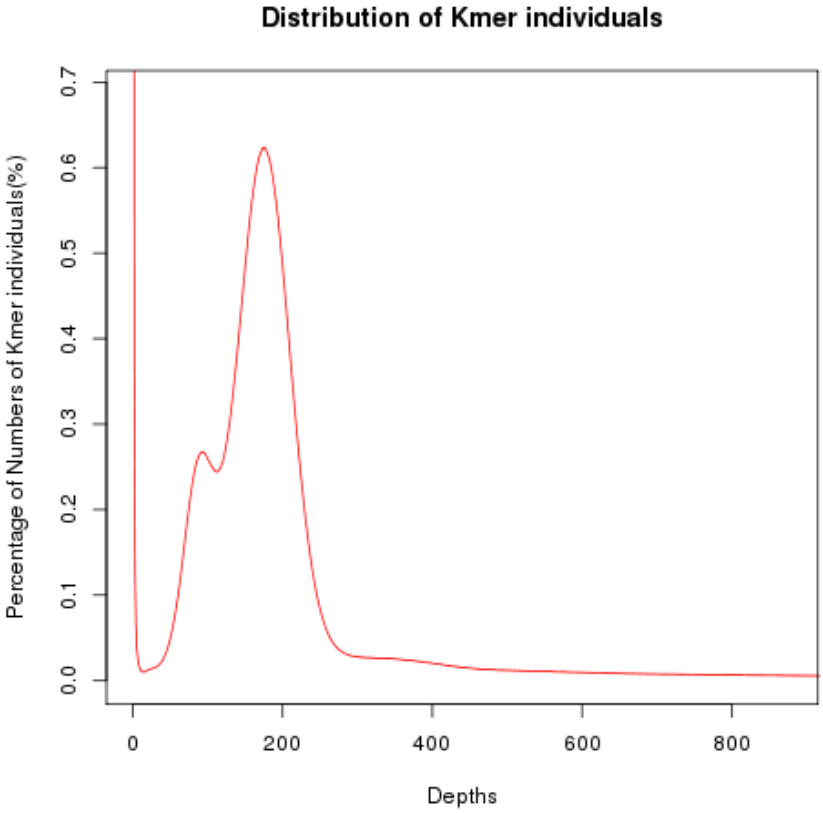
TABLE 2 Statistics of repeat elements of *Caligula japonica.*

Repeat types	**Number of elements**	**Length occupied(bp)**
SINEs	38233	6217105
LINEs	229663	75413600
LTR elements	21732	22079242
DNA transposons	112315	41673523
Unclassified	716097	119312633
Small RNA	1156	254558
Simple repeats	94116	4155387
Low complexity	14742	690658
Bases masked		342125543

Species	DH-PBAN	PTTH	cry1	cry2	per	tim
Drosophila melanogaster	0	0	1	0	1	1
Danaus plexippus	0	0	1	1	1	1
Helicoverpa armigera	1	1	1	1	1	1
Spodoptera frugiperda	1	0	2	1	1	0
Bombyx mori	1	1	1	1	1	1
Samia ricini	1	1	1	1	0	1
Caligula japonica	1	1	1	1	0	1
Antheraea pernyi	1	1	1	1	1	0
Antheraea yamamai	1	0	1	1	0	0
Plutella xylostella	0	0	1	1	0	1

TABLE 3 The number of diapause-related genes in *Caligula japonica.*

FIGURE 1 Morphological photographs of *Caligula japonica* . (a) Cocoon containing pupae (b) Female adult. (c) Male adult.



Hosted file

`image5.emf` available at https://authorea.com/users/520086/articles/593538-a-chromosome-level-genome-supplies-an-insight-into-the-distinctive-diapause-characteristics-in-caligula-japonica

FIGURE 2 (a) Distribution of K-mer frequency in *Caligula japonica* genome sequencing reads at different depths. (b) Chromosome Hi-C interaction map of *Caligula japonica* identified 31 linkage groups.

Hosted file

`image6.emf` available at https://authorea.com/users/520086/articles/593538-a-chromosome-level-genome-supplies-an-insight-into-the-distinctive-diapause-characteristics-in-caligula-japonica

FIGURE 3 Circos graph of characteristics of *Caligula japonica* genome. Tracks a-d represent karyotype, count of gene, repeats density and GC density in the genome of *C japonica* .

Hosted file

`image7.emf` available at https://authorea.com/users/520086/articles/593538-a-chromosome-level-genome-supplies-an-insight-into-the-distinctive-diapause-characteristics-in-caligula-japonica

FIGURE 4 Phylogenetic tree and orthologs between genomes of*Caligula japonica* and 9 other insects. The maximum likelihood phylogenomic tree was calculated based on 1167 single-copy genes. ("1:1:1": single-

copy universal genes; “N:N:N”: other multiply genes; “Species-species OGs”: genes without in any other species; “Unassigned genes”: genes not assigned to any homologous group. ; “SD”: specific duplicated genes ; “Patchy”: orthologous in all other species;).