



OPEN

DATA DESCRIPTOR

Chromosome-scale genome assembly of the high royal jelly-producing honeybees

Lianfei Cao¹✉, Xiaomeng Zhao², Yanping Chen³ & Cheng Sun²✉

A high royal jelly-producing strain of honeybees (HRJHB) has been obtained by successive artificial selection of Italian honeybees (*Apis mellifera ligustica*) in China. The HRJHB can produce amounts of royal jelly that are dozens of times greater than their original counterparts, which has promoted China to be the largest producer of royal jelly in the world. In this study, we generated a chromosome-scale of the genome sequence for the HRJHB using PacBio long reads and Hi-C technique. The genome consists of 16 pseudo-chromosomes that contain 222 Mb of sequence, with a scaffold N50 of 13.6 Mb. BUSCO analysis yielded a completeness score of 99.3%. The genome has 12,288 predicted protein-coding genes and a rate of 8.11% of repetitive sequences. One chromosome inversion was identified between the HRJHB and the closely related Italian honeybees through whole-genome alignment analysis. The HRJHB's genome sequence will be an important resource for understanding the genetic basis of high levels of royal jelly production, which may also shed light on the evolution of domesticated insects.

Background & Summary

Royal jelly (RJ) is a proteinaceous secretion synthesized by the hypopharyngeal and mandibular glands of nurse worker bees and is used for feeding queen and larvae¹. It also plays a critical role in the caste determination of honeybees². Nowadays, RJ is widely used in medical products, health foods and cosmetics in many countries owing to the numerous biological activities it is known to perform including anti-bacterial, anti-oxidative, anti-inflammatory, immunomodulatory, anti-tumoral, and anti-aging activities^{3,4}. China is now the largest producer and exporter of RJ in the world, which satisfies nearly all the global demand⁵. Since the 1980s, the yearly production of RJ in China has increased from 200 to around 3000 tons⁵. The rapidly increased production of RJ in China has been mainly attributed to the successful breeding of the high royal jelly-producing honeybees (HRJHB) (Fig. 1), and the effective utilization of corresponding production tools and techniques⁶.

HRJHB was derived from an Italian honeybee subspecies (*Apis mellifera ligustica*), which was chiefly introduced into China in the 1910s–1930s⁷. In 1960s, attempts were made by beekeepers in the Southeast region of China to select high RJ producing bee stocks to meet a high demand for RJ⁸. The colony that displayed a high rate of RJ production was selected for raising daughter queens and drones in each apiary⁸. Sometimes queens were also developed using larvae of high RJ producing colonies from different apiaries⁸. Queens then open-mated with local drones in the air⁸. After the aforementioned semi-controlled style of breeding, the annual RJ production per colony increased from 0.2–0.3 kg in the 1960s to 2–3 kg in the late 1980s and even reaching 6–8 kg in the 2000s⁸. This was perceived as a miracle and the HRJHB was rapidly introduced to other regions of China from the 1980s, onwards as well as other countries at a later date. At present, the annual production per HRJHB colony has reached more than 10 kg, which is dozens of times greater than that of common Italian honeybees (*A. m. ligustica*)⁵. RJ production has become a major income source for many beekeepers in China and the HRJHB has been certified as a new honeybee genetic resource by the Chinese government⁷.

Previous studies regarding isoenzymes, microsatellites and mitochondrial DNA have shown significant genetic differentiation between the HRJHB and the other common *A. m. ligustica* populations in China^{9–11}. It was suggested that morphological markers, behavioural and physiological changes, and differently expressed proteins and genes, correlate to the high royal jelly-producing trait^{12–16}. However, related research has so failed to

¹Institute of Animal Husbandry and Veterinary Science, Zhejiang Academy of Agricultural Sciences, Hangzhou, 310021, China. ²Institute of Apicultural Research, Chinese Academy of Agricultural Sciences, Beijing, 100093, China.

³USDA-ARS Bee Research Laboratory, USDA-ARS, Bldg. 306, BARC-East, Beltsville, MD, 20705, USA. ✉e-mail: caolf@mail.zaas.ac.cn; suncheng@caas.cn

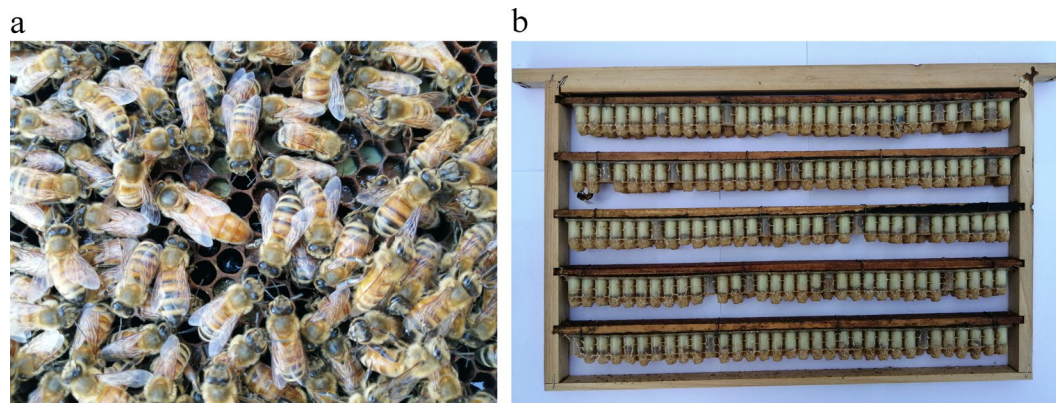


Fig. 1 High royal jelly-producing honeybees (HRJHB) in China. **(a)** Queen and workers in one colony. **(b)** Royal jelly in the queen's cells.

Genome sequencing			
	Read number	Read_length(mean)	Total read length (Gb)
PacBio long reads	2,154,163	15,489	33.37
Illumina sequencing	71786450	150	10.77
RNA-seq	130258000	150	18.05
Hi-C sequencing	218592996	150	32.79
Genome assembly			
Genome assembly size	222 Mb		
Number of scaffolds	16		
Scaffold N50	13.6 Mb		
BUSCO completeness	99.30%		

Table 1. Sequencing data generated for the HRJHB genome assembly.

develop an entirely clear picture of what causes the complex royal jelly-producing trait. In recent years, honeybee selection programs for high RJ production have also been implemented in Brazil and France beekeeping^{17,18}. Additionally, further breeding of HRJHB for improving general resistance to disease is being carried out in China.

In this study, we generated a chromosome-scale of the genome assembly of the HRJHB using PacBio long-reads, Illumina short reads, and the Hi-C chromosome conformation capture technique (Table 1; Fig. 2a). The resultant genome has a total length of 222 Mb with 16 chromosomes, and the scaffold N50 was 13.6 Mb (Table 1). One chromosome inversion was identified between HRJHB and the closely related Italian honeybees via whole-genome alignment analysis (Fig. 2b). Moreover, through a combination of *ab initio* gene predictions, transcript evidence and homologous protein evidence, 12,288 protein coding genes were identified in this genome, therein 6,615 genes were assigned a GO term and 8,614 genes were assigned a protein domain (Table 2). Repetitive elements are made of 8.11% of the HRJHB genome sequence, but transposable elements (TEs) only occupy 2.15% (Table 2). Among those TEs, DNA transposons represented the most abundant TE class, which make up the majority of the total TE content (1.68% out of 2.15%). Furthermore, *Tc1*-mariner (TcMar) is the most abundant TE superfamily in the genome. The genome sequence provides a valuable resource for exploring the molecular basis of the high royal jelly-producing trait in honeybee and will facilitate further genetic improvements. The HRJHB may even represent a novel animal model for studying the effects of artificial selection on insects.

Methods

Sample collection and genome sequencing. Samples of the HRJHB for genome and transcriptome sequencing were collected in 2019 from Zhejiang Province, China, where the HRJHB was originated and primarily distributed (Fig. 3).

Newly emerged drone bees ($n = 20$), that are descendants of the queen bee, were collected from a single colony (Fig. 1a). The thoraxes were pooled for PacBio single molecule real-time (SMRT) sequencing and Illumina HiSeq sequencing. Genomic DNA was extracted using the Genra Puregene Tissue Kit (Qiagen) and was sequenced in accordance with the standard protocols. Newly emerged worker bees ($n = 20$) were collected from the same colony and their thoraxes were pooled for Hi-C sequencing. Hi-C library preparation was performed by Frasiergen (<http://www.frasiergen.com/>), which mainly followed a protocol described previously¹⁹. The obtained Hi-C sequencing libraries were sequenced on the Illumina HiSeq X Ten platform. Worker bees that

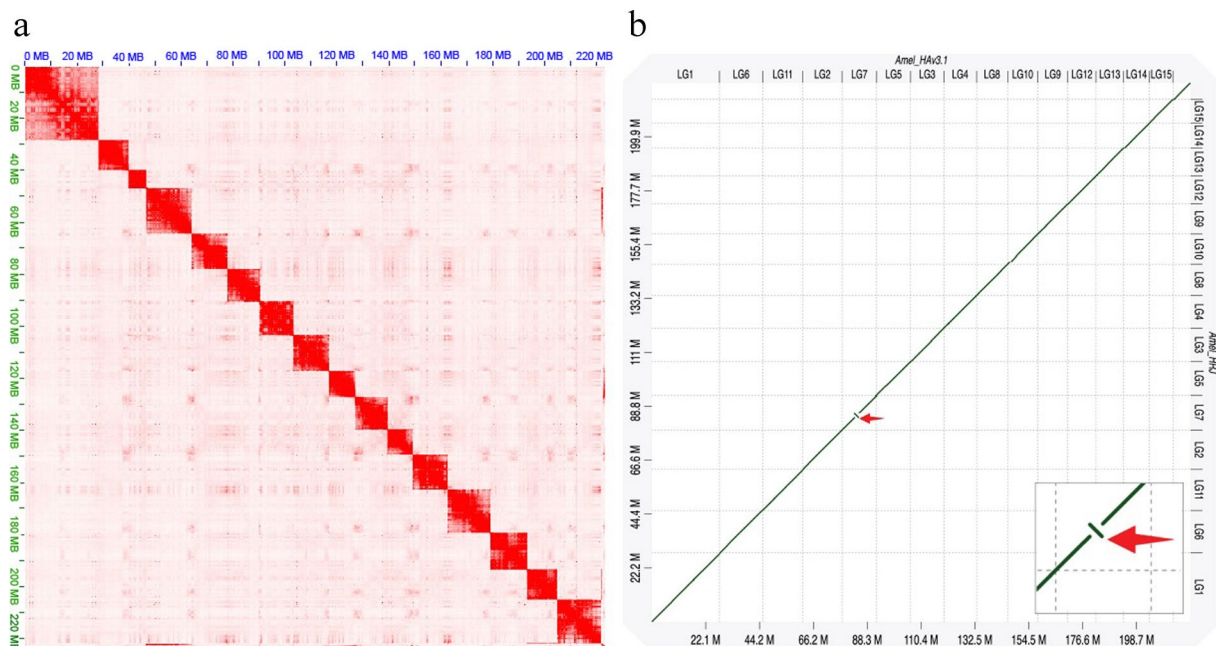


Fig. 2 Chromosome-scale assembly for HRJHB genome. **(a)** The HRJHB's genome contig contact matrix using Hi-C data. **(b)** The HRJHB's genome sequence was aligned with a closely related honeybee genome (NCBI assembly: Amel_HAv3). The red arrow indicates the chromosome inversion between the two genomes on LG7.

Protein-coding genes			
	Total gene number	12,288	
	BUSCO completeness	97%	
	Number of genes with a GO term	6,615	
	Number of genes with a protein domain	8,614	
Repetitive sequences			
	TE superfamily	Length occupied (bp)	Percent of genome
DNA transposons	TcMar	3557056	1.67
	hAT	14338	0.01
non-LTR retrotransposons	CR1	952652	0.45
	R2	49376	0.02
LTR retrotransposons	Copia	6014	0.00
	Gypsy	383280	0.18
Total TEs		4962716	2.15
Other repeats		13761592	5.96
Total repeats		18724308	8.11

Table 2. Annotation of protein-coding genes and repetitive sequences.

were excreting royal jelly ($n = 20$) were collected from the same colony and their heads, thoraxes and abdomens (excluding the mid-gut tissues) were pooled for RNA-seq on the Illumina HiSeq X Ten platform.

De novo genome assembly for HRJHB. A total of 33.37 Gb of long reads were generated by the PacBio Sequel platform (Table 1), which were self-corrected and assembled into contigs using Canu v2.1²⁰, with default parameters. The obtained contigs were parsed by Purge Haplotigs v1.1.1²¹ to get rid of the redundancies caused by the heterozygosity of the pooled honeybee samples. Then, the remaining non-redundant contigs were polished with Illumina HiSeq reads (Table 1) three times by utilizing software Pilon v1.23²². Finally, the Juicer tool²³ was applied to map Hi-C reads (Table 1) against the polished contig sequences of HRJHB using the BWA algorithm²⁴. The 3D-DNA pipeline²⁵ was applied to scaffold the contig sequences in relation to the chromosome-scale of genome assembly.

Annotation of repeat sequences. TEs were *de novo* identified by RepeatModeler2²⁶, in line with default parameters. Using the obtained repeat library, each honeybee genome assembly was analyzed with RepeatMasker (<http://www.repeatmasker.org>) to yield a comprehensive summary of the TE landscape in each assembly. The

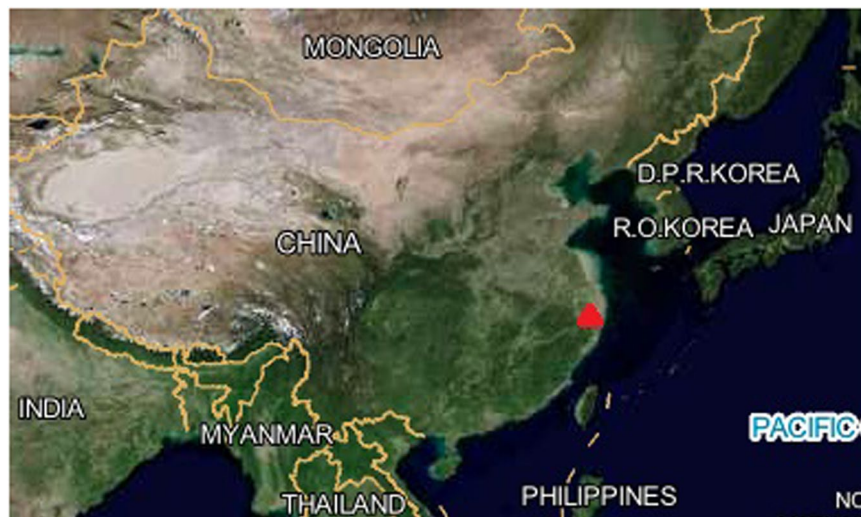


Fig. 3 Original area of HRJHB (red arrowhead).

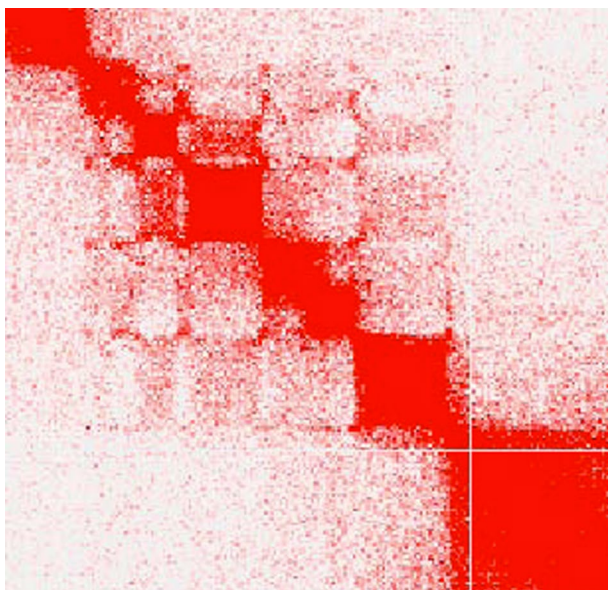


Fig. 4 Hi-C heatmap around the identified chromosome inversion region in the HRJHB.

annotation files produced by RepeatMasker were processed by in-house scripts to eliminate redundancies. In addition, refined annotation files were used to determine the TE diversity and abundance within each assembly and tandem repeats were identified with the Tandem Repeat Finder²⁷, which was implemented in RepeatMasker.

Prediction and functional annotation of protein-coding genes. Annotation of protein-coding genes was based on *ab initio* gene predictions, transcript evidence, and homologous protein evidence, which were all applied in the MAKER computational pipeline²⁸. Meanwhile, RNA-seq reads obtained in this study were assembled using Trinity²⁹. The assembled RNA-seq transcripts, along with proteins from bees (superfamily Apoidea) that are available in the National Center for Biotechnology Information (NCBI) GenBank (last accessed on 01/28/2020), were imported into the MAKER pipeline to generate gene models. To obtain functional clues for the predicted gene models, protein sequences encoded by them were searched against the Uniprot-Swiss-Prot protein databases (last accessed on 01/28/2020) using the BLASTp algorithm implemented in BLAST suite v2.28³⁰. In addition, protein domains and GO terms associated with gene models were identified by InterproScan-5³¹.

Data Records

The raw data was submitted to the National Center for Biotechnology Information (NCBI) SRA database (Experiments for SRP300170) under BioProject accession number PRJNA689474³². The assembled genome has been deposited at DDBJ/ENA/GenBank under the accession GCA_019321825.1³³. Moreover, the genome annotation results have been deposited at the Figshare database³⁴.

Technical Validation

Evaluation of the genome assembly. The completeness of the genome assembly was evaluated using a set of 4,415 hymenopteran benchmarking universal single-copy orthologs (BUSCOs) using software BUSCO v3³⁵. The results indicated that 99.3% of these BUSCOs were present in the genome assembly (Table 1), suggesting a remarkably complete assembly of the HRJHB genome.

Furthermore, the chromosome-level structural accuracy was assessed by performing whole-genome alignments between HRJHB genome and a closely related honeybee genome (GenBank assembly: Amel_HAv3) using software D-GENIES³⁶. The alignment results revealed a highly conserved chromosome structure between the two genomes, indicating an accurate scaffolding of contigs in the HRJHB genome. Nevertheless, we did find one inversion on LG7 (Fig. 2b). The Hi-C heatmap revealed a well-organized interaction contact pattern along the diagonals within/around the chromosome inversion region in HRJHB (Fig. 4), which rules out the possibility that the structural variation was derived from unreliable Hi-C signals in the HRJHB assembly. In addition, as chromosome inversion has been found to be associated with honeybee adaptations³⁷, the inversion identified in the HRJHB genome will guarantee that further analysis will be carried out to investigate its association with high royal jelly production.

Code availability

All software used in this work is in the public domain, with parameters being clearly described in Methods. If no detail parameters were mentioned for a software, default parameters were used as suggested by developer.

Received: 9 August 2021; Accepted: 3 November 2021;

Published online: 25 November 2021

References

- Knecht, D. & Kaatz, H. H. Patterns of larval food production by hypopharyngeal glands in adult worker honey bees. *Apidologie* **21**, 457–468, <https://doi.org/10.1051/apido:19900507> (1990).
- Kamakura, M. Royalactin induces queen differentiation in honeybees. *Nature* **473**, 478–483, <https://doi.org/10.1038/nature10093> (2011).
- Ramadan, M. F. & Al-Ghamdi, A. Bioactive compounds and health-promoting properties of royal jelly: A review. *J Funct Foods* **4**, 39–52, <https://doi.org/10.1016/j.jff.2011.12.007> (2012).
- You, M. M. *et al.* Royal jelly alleviates cognitive deficits and b-amyloid accumulation in APP/PS1 mouse model via activation of the cAMP/PKA/CREB/BDNF pathway and inhibition of neuronal apoptosis. *Front Aging Neurosci* **10**, 428, <https://doi.org/10.3389/fnagi.2018.00428> (2019).
- Zheng, H. Q., Cao, L. F., Huang, S. K., Neumann, P. & Hu, F. L. Current status of the beekeeping industry in China, In: Chantawannakul P., Williams G., Neumann P. (eds) *Asian Beekeeping in the 21st Century*. Springer Nature Singapore Pte Ltd., Singapore, 129–158 (2018).
- Hu, F. L. *et al.* Standard methods for *Apis mellifera* royal jelly research. *J Apic Sci* **58**, 1–68, <https://doi.org/10.1080/00218839.2017.1286003> (2019).
- (CNCAGR) China National Commission of Animal Genetic Resources. *Animal genetic resources in China –Bees*. Chinese Agricultural Press, Beijing, China (2011).
- Cao, L. F., Zheng, H. Q., Pirk, C. W. W., Hu, F. L. & Xu, Z. W. High royal jelly-producing honey bees (*Apis mellifera ligustica*) (Hymenoptera: Apidae) in China. *J Econ Entomol* **109**, 510–514, <https://doi.org/10.1093/jee/tow013> (2016).
- Sun, L. X., Chen, Z. Y., Yuan, J. J. & Xie, J. J. Genetic variability of MDHIII in four lines of *Apis mellifera ligustica*. *J Zhangzhou Teach Coll* **17**, 54–59 (2004).
- Chen, S. L., Li, J. K., Zhong, B. X. & Su, S. K. Microsatellite analysis of royal jelly producing traits of Italian honeybee (*Apis mellifera ligustica*). *Acta Genet Sin* **32**, 1037–1044 (2005).
- Cao, L. F., Zheng, H. Q., Shu, Q. Y., Hu, F. L. & Xu, Z. W. Mitochondrial DNA characterization of high royal jelly-producing honeybees (Hymenoptera: Apidae) in China. *J Apic Sci* **61**, 217–222, <https://doi.org/10.1515/jas-2017-0016> (2017).
- Li, J. K., Feng, M., Desalegn, B., Fang, Y. & Zheng, A. J. Proteome comparison of hypopharyngeal gland development between Italian and royal jelly producing worker honeybees (*Apis mellifera* L.). *J Proteome Res* **9**, 6578–6594, <https://doi.org/10.1021/pr100768t> (2010).
- Wu, F. *et al.* Behavioural, physiological and molecular changes in alloparental caregivers may be responsible for selection response for female reproductive investment in honey bees. *Mol Ecol* **28**, 4212–4227, <https://doi.org/10.1111/mec.15207> (2019).
- Altaye, S. Z., Meng, L. F. & Li, J. K. Molecular insights into the enhanced performance of royal jelly secretion by a stock of honeybee (*Apis mellifera ligustica*) selected for increasing royal jelly production. *Apidologie* **50**, 436–453, <https://doi.org/10.1007/s13592-019-00656-1> (2019).
- Nie, H. Y. *et al.* Identification of genes related to high royal jelly production in the honey bee (*Apis mellifera*) using microarray analysis. *Genet Mol Biol* **789**, 781–789, <https://doi.org/10.1590/1678-4685-GMB-2017-0013> (2017).
- Rizwan, M. *et al.* Population genomics of honey bees reveals a selection signature indispensable for royal jelly production. *Mol Cell Probes* **52**, 101542, <https://doi.org/10.1016/j.mcp.2020.101542> (2020).
- Parpinelli, R. S., Ruvolo-Takasusuki, M. C. C. & Toledo, V. A. A. MRJP microsatellite markers in Africanized *Apis mellifera* colonies selected on the basis of royal jelly production. *Genet Mol Res* **13**, 6724–6733, <https://doi.org/10.4238/2014.August.28.16> (2014).
- Wragg, D. *et al.* Whole-genome resequencing of honeybee drones to detect genomic selection in a population managed for royal jelly. *Sci Rep* **6**, 27168, <https://doi.org/10.1038/srep27168> (2016).
- Belton, J. M. *et al.* Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276, <https://doi.org/10.1016/j.ymeth.2012.05.001> (2012).
- Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **27**, 722–736, <https://doi.org/10.1101/gr.215087.116> (2017).
- Roach, M. J., Schmidt, S. A. & Borneman, A. R. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**, 460, <https://doi.org/10.1186/s12859-018-2485-7> (2018).
- Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963, <https://doi.org/10.1371/journal.pone.0112963> (2014).
- Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95–98, <https://doi.org/10.1016/j.cels.2016.07.002> (2016).
- Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595, <https://doi.org/10.1093/bioinformatics/btp698> (2010).

25. Dudchenko, O. *et al.* *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95, <https://doi.org/10.1126/science.aal3327> (2017).
26. Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci USA* **117**, 9451–9457, <https://doi.org/10.1073/pnas.1921046117> (2020).
27. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**, 573–580, <https://doi.org/10.1093/nar/27.2.573> (1999).
28. Cantarel, B. L. *et al.* MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* **18**, 188–196, <https://doi.org/10.1101/gr.6743907> (2008).
29. Haas, B. J. *et al.* *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**, 1494–1512, <https://doi.org/10.1038/nprot.2013.084> (2013).
30. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402, <https://doi.org/10.1093/nar/25.17.3389> (1997).
31. Jones, P. H. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240, <https://doi.org/10.1093/bioinformatics/btu031> (2014).
32. NCBI Sequence Read Archive <https://identifiers.org/insdc.sra:SRP300170> (2021).
33. NCBI Assembly https://identifiers.org/ncbi/insdc.gca:GCA_019321825.1 (2021).
34. Sun, C. Genome annotation for high royal jelly-producing honeybee. *figshare* <https://doi.org/10.6084/m9.figshare.14702193> (2021).
35. Waterhouse, R. M. *et al.* BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol* **35**, 543–548, <https://doi.org/10.1093/molbev/msx319> (2018).
36. Cabanettes, F. & Klopp, C. D-GENIES: Dot plot large genomes in an interactive, efficient and simple way. *PeerJ* **6**, e4958, <https://doi.org/10.7717/peerj.4958> (2018).
37. Christmas, M. J. *et al.* Chromosomal inversions associated with environmental adaptation in honeybees. *Mol Ecol* **28**, 1358–1374, <https://doi.org/10.1111/mec.14944> (2019).

Acknowledgements

This work was funded by the Science and Technology Department of Zhejiang Province, China (2016C02054-11), the National Natural Science Foundation of China (31602014), and the Fundamental Research Funds of Chinese Academy of Agricultural Sciences (grant numbers: Y2019XK13, Y2021XK16).

Author contributions

L.C. and C.S. conceived the study. L.C. collected the samples. L.C. extracted the genomic DNA and conducted sequencing. C.S. and X.Z. performed bioinformatics analysis. C.S., L.C. and Y.C. wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to L.C. or C.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2021