scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Chromosome-level genome assembly of an endoparasitoid *Cotesia ruficrus*

Xianxin Zhao[1], Yuanyuan Liu[1], Xinhai Ye[2,3], Fang Wang[1], Qi Fang[1] & Gongyin Ye[1]✉

Cotesia ruficrus **is a gregarious larval endoparasitoid of many important agricultural pest moths, such as the fall armyworm, northern armyworm and rice leaffolder. While eight of *Cotesia* genomes had been reported, high-quality genomic data of *C. ruficrus* has yet to be performed. This study presents a chromosome-scale genome assembly of *C. ruficrus*, utilizing both PacBio HiFi-reads and Hi-C reads. The genome size is about 185.3 Mb harboring 10 chromosomes, and 209 unanchored scaffolds with a scaffold N50 length of 15.81 Mb. The anchored rate of sequences to chromosomes is 95.55%. A total of 14,001 genes have been annotated. This work delivers a first high-quality, chromosome-scale genome for *C. ruficrus*, enriching the datasets for comparative evolutionary studies within the *Cotesia* genus. Additionally, it offers a valuable genomic resource for future biological and genetic research, as well as for enhancing pest management strategies, such as biological control.**

Background & Summary

Cotesia ruficrus (Haliday) previously known as *Apanteles ruficrus* Haliday, is a member of the Braconidae family within the Ichneumonoidea superfamily (Hymenoptera)[1–6]. This endoparasitoid wasp completes its larval development by utilizing finite resources within a host. This species is known to parasitize a variety of moth species and has an extensive geographical distribution across the Afrotropical, Australasian, Neotropical, Oriental, and Palaearctic regions[4,5,7]. Notably, *C. ruficrus* demonstrates significant potential in controlling *Spodoptera frugiperda*[3,5], an invasive pest originating from the tropical and subtropical Americas that feeds on over than 300 plant species including maize, rice, barley, cotton, soybean and vegetables and has spread to much of Africa and Asian[8–11]. Besides *S. frugiperda*, other key agricultural pests such as *Cnaphalocrocis medinalis*[4,12], *Mythimna separata*[2,6], *Heliothis armigera*[13,14], *Naranga aenescens*[15,16] and *Chilo suppressalis*[17,18], are also hosts to *C. ruficrus*. As the importance of *C. ruficrus* as a biological control agent against numerous crop pests becomes more recognized[4], research into its biology has grown. Studies have delved into its mechanisms of host selection and location involving olfactory processes[19,20], the impact of parasitism on host development[4,5,21], and assessments of its potential for biological control and host preferences[13,18]. Despite the increasing acknowledgment of *Cotesia ruficrus*'s role as a vital parasitoid for integrated pest management, there remains a considerable lack of detailed genomic information, particularly concerning its complete genome sequence and the functional genes associated with host recognition and parasitism. Enhancing our understanding in these areas could further enhance its application in agricultural pest control.

Genomic data is crucial for delving into genetic and evolutionary studies. HiFi reads facilitate the *de novo* assembly of genomes with remarkable completeness, contiguity, and accuracy[22,23]. Furthermore, Hi-C technology has played a pivotal role in achieving chromosome-level genome assemblies[24]. In this study, we employed an integrative approach, utilizing both PacBio HiFi and Hi-C data, to assemble the genome of *C. ruficrus*. We obtained approximately 13 Gb of high-quality PacBio HiFi reads with an average length of 16,679 bp. Based on k-mer analysis, the estimated genome size is approximately 172.9 Mb, and heterozygosity rate is about 0.16% (Fig. 1A). The final assembled genome for *C. ruficrus* was 185.3 Mb, with 95.55% of sequences anchored. A total of 79.2 Mb of sequences were identified as repetitive elements, accounting for approximately 42.73% of the entire

[1]State Key Laboratory of Rice Biology and Breeding & Ministry of Agricultural and Rural Affairs Key Laboratory of Molecular Biology of Crop Pathogens and Insects, Institute of Insect Sciences, Zhejiang University, Hangzhou, China. [2]College of Advanced Agriculture Sciences, Zhejiang A&F University, Hangzhou, China. [3]Zhejiang Key Laboratory of Biology and Ecological Regulation of Crop Pathogens and Insects, Zhejiang A&F University, Hangzhou, 311300, China. ✉e-mail: chu@zju.edu.cn

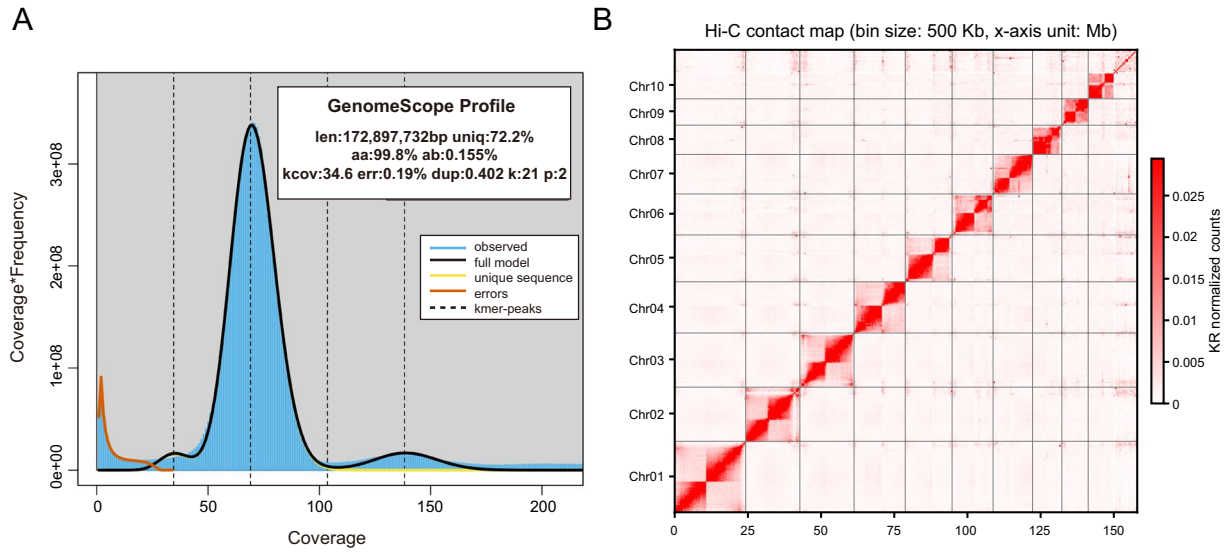

Fig. 1 Genome Survey of *Cotesia ruficrus* and Chromosomes Assembly. (**A**) Kmer-21-spectra output generated by GenomeScope2 using PacBio HiFi reads. The genome size and heterozygosity rate were also estimated. (**B**) The Hi-C heatmap showing chromosome interactions among the chromosomes of *C. ruficrus*.

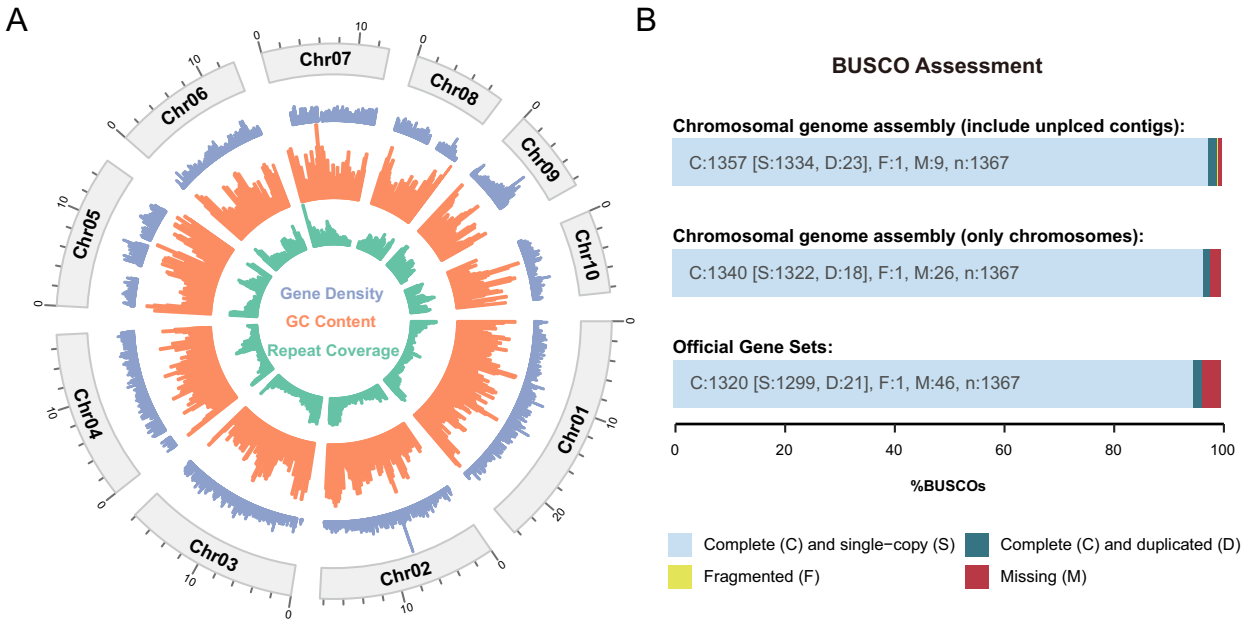

Fig. 2 Characteristics of *Cotesia ruficrus* Genome Assembly. (**A**) Overview of *C. ruficrus* chromosomes: Gene density (purple), GC content (orange) and repeat coverage (green) are plotted. (**B**) BUSCO scores (insect_odb10) for the assembled genome and protein-coding genes of *C. ruficrus*.

genome (exclude Ns). Additionally, we conducted gene annotation on the genome, identifying 14,001 genes. The assessment of genome completeness and continuity using BUSCO showed highly-quality results, with a completeness score of 99.27% (Fig. 2A). Here, we generated a high-quality, chromosome-level genome assembly for *C. ruficrus*, which serves as a valuable resource for deepening our understanding of the genetic mechanisms and evolutionary drivers that facilitate adaptation to diverse hosts and provide a foundation for developing more effective and sustainable biological control strategies against major agricultural pests. Currently, eight *Cotesia* genomes are publicly available[25,26]. Furthermore, chromosome-level assembly of *C. ruficrus* will aid in clarifying the evolutionary relationships within the *Cotesia* genus, as well as in identifying shared or unique traits (e.g., wasps and their endogenous polydnaviruses; host preferences) and behaviors through comparative genomics analyses[27,28].

Sequencing Category	Reads Type	Sequencing Platform	Reads Bases (Gb)
DNA sequencing	PacBio HiFi reads	PACBIO_SMRT (Sequel II)	13.04 Gb
	Hi-C illumina reads	Illumina HiSeq 2000	17.32 Gb
RNA sequencing	RNA-seq reads	Illumina NovaSeq 5000	64.8 Gb

Table 1. Statistics of DNA and RNA sequencing.

Methods

Wasps rearing and DNA extraction. *C. ruficrus* were sampled in Zhejiang and reared using 10% hydromel at 25 ± 1°C and 75 ± 5% relative humidity, with a photoperiod of 14 L:10D[4]. Adult samples were collected for DNA extraction. To prepare high molecular weight genomic DNA, we employed the SDS method, followed by purification using the QIAGEN® Genomic kit (Cat#13343, QIAGEN). This process adhered to the standard operating procedures outlined by the manufacturer.

RNA sequencing and quality control. Total RNA was extracted from three adults of each sex. A total of 1 μg RNA per sample was used as input material for the RNA sample preparations. RNA quality and quantity were assessed using the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA) and Qubit® RNA Assay Kit with a Qubit® 2.0 Flurometer (Life Technologies, CA, USA), respectively. Sequencing libraries were generated using TruSeq RNA Library Preparation Kit (Illumina, USA) according to the manufacturer's recommendations, and index codes were added to attribute sequences to each sample. The validated library preparations were then sequenced on an Illumina NovaSeq platform after cluster generation of index-coded samples. To ensure the reliability and accuracy of our RNA-seq data, we performed several quality control steps before proceeding with downstream analyses, including reads quality assessment with FastQC, reads trimming and filtering with Trimmomatic, and get the final high-quality reads[29,30] by aligning trimmed reads to the reference genome using STAR 2.7.10b[31].

HiFi SMRTbell library construction and sequencing. SMRTbell target size libraries were prepared for sequencing using PacBio's protocol (Pacific Biosciences, CA, USA) with 10 kb or 20 kb solutions. Key steps included gDNA shearing, DNA damage repair, blunt-end ligation with SMRTbell adapters, size selection, and polymerase binding. A total of 2 μg DNA was used for the DNA library preparation. Sequencing was conducted on a PacBio Sequel II system using the Sequel II Kit 2.0 at Nextomics. Raw sequencing data (also known as raw polymerase reads) were processed using the quality control protocol in SMRTlink v7.0 to remove low-quality reads and adapters, resulting in subreads. Total of 782,091 circular consensus sequence (CCS) reads[32], accounting for 13,044,522,524 bp passed quality control using ccs tool (parameters: "--min-passes 1 --min-rq 0.99 --min-length 100") (https://github.com/PacificBiosciences/ccs) (Table 1).

Hi-C library construction and sequencing. To anchor scaffolds to chromosomes, genomic DNA from *C. ruficrus* was extracted for Hi-C library construction and sequenced on the Illumina Novaseq platform. The process began by grinding the entire bodies of *C. ruficrus* in a nuclei isolation buffer containing 2% formaldehyde. The crosslinking reaction was halted with glycine, followed by vacuum infiltration. The fixed tissue was ground into a powder and re-suspended in nuclei isolation buffer to yield a nuclei suspension. Purified nuclei were digested with 100 units of DpnII and labeled by incubation with biotin-14-dCTP. Unligated DNA ends were processed with T4 DNA polymerase to remove biotin-14-dCTP. The ligated DNA was then sheared into 300~600 bp fragments, subjected to blunt-end repair and A-tailing, and purified using biotin-streptavidin pull-down. The Hi-C libraries were subsequently quantified and sequenced using the Illumina Novaseq platform. A total of 17,323,597,694 bp clean reads[33] were generated after quality control using fastp v0.19.4[34] with default settings (Table 1).

Genome size estimation. To estimate the genome size of *C. ruficrus*, a k-mer analysis was performed using HiFi reads. Using GenomeScope[35], the frequency of 17-mers was determined. The formula used to estimate the genome size is: G = K_num/K_depth, where G represents the genome size, K_depth is the depth of k-mers, and K_num denotes the total count of 17-mers. The genome size was estimated to be approximately 172.9 Mb (Fig. 1), which is similar to the genome sizes (165.9~288.9 Mb) of other publicly reported *Cotesia* species[25,26].

Genome assembly. The genome was *de novo* assembled into contigs using NextDenovo v2.5.0[36] software in "hifi" mode, which involved the following steps: (1) overlapping raw subreads for error correction; (2) preassembly and error correction of the reads; (3) detecting overlaps in the error-corrected reads and filtering these overlaps; (4) constructing a graph from the overlaps and generating contigs from the graph.

All PacBio SMRT reads were then realigned to the contigs using Blasr tool[37] (https://github.com/PacificBiosciences/blasr) and Arrow algorithm was employed to correct the sequencing errors based on the alignments with default parameters. To further improve the accuracy of the assembly, NextPolish v1.2.4[38] was used to refine the contigs with Illumina short reads, also with default settings. To discard potentially redundant contigs and generate a final assembly, similarity searches were performed with the parameters "identity 0.8 –overlap 0.8". To exclude mitochondrial sequences from the assembly, the draft genome was compared against the NT library, and any aligning sequences were removed. A total of 185,307,737 bp were assembled in the final assembly (Table 2)[39].

	Preliminary Assembly	Chromosome-level Genome Assembly
Number of contigs/scaffolds	281	84
Contig/Scaffold N50 (bp)	5,000,615	15,811,414
Longest contigs/scaffolds (bp)	13,257,128	24,362,489
Size of final assembly (bp)	185,307,737	185,315,137

Table 2. Statistics of genome assembly.

Chromosome ID	Sequences Length	Anchored Scaffold Number
Chr01	24,362,489	3
Chr02	18,382,948	5
Chr03	18,126,106	10
Chr04	17,412,739	11
Chr05	15,811,414	11
Chr06	13,517,456	8
Chr07	13,117,360	11
Chr08	9,606,537	8
Chr09	8,645,650	7
Chr10	8,539,638	10
Total	147,522,337	84

Table 3. Statistics for chromosomes sequence length.

Hi-C scaffolding using Hi-C reads. A total of 117.15 million raw paired-end reads were obtained from the sequencing libraries. The quality control of the Hi-C raw data was then performed. Initially, sequences of low quality (with quality scores ≤ 15), adapter sequences, and those shorter than 30 base pairs were removed using fastp v0.19.4[40]. This resulted in 115.75 million clean paired-end reads, totaling 17,323,597,694 bp (Table 1). Following this, the cleaned paired-end reads were aligned to the preliminary assembly using Bowtie2 v2.3.2[41] with parameters set for "-end-to-end --very-sensitive -L 30". This process yielded 9,221,219 uniquely mapped paired-end reads, accounting for 15.93% of all clean reads. The raw FASTQ files were subsequently processed and analyzed with Juicer v1.6[42], and the output was used to create chromosomal-length scaffolds through 3D-DNA (pipeline version: 180419)[43], employing several default procedures. Any placement and orientation errors that showed clear discrete chromatin interaction patterns were manually corrected and visualized using Juicebox v2.20.00[44]. These corrections were also supported by other reports on karyotype or chromosomes in *Cotesia* species ($n = 10$ both in *C. chilonis* and *C. glomerata*)[26,45] and Microgastrinae subfamily ($n = 9$~11)[46]. A total of 84 scaffolds were assembled into 10 chromosomes (Fig. 1B; Table 3).

Repetitive element annotation. Repeats were masked using RepeatMasker v4.1.6[47] and RepeatModeler v2.0.4[48]. Briefly, the options "-a" and "-d" were first added to include ancestors and descendants of Insecta using the FamDB tool to build a reference repeat database. Additionally, RepeatModeler v2.0.4, which includes both RECON v1.08[49] and RepeatScout v1.0.6[50], was used to identify *de novo* repeats. Results were further imported into RepeatMasker for final prediction with rmblastn v2.13.0+[51]. Through *de novo*-based and homology-based prediction, the genome was found to have a total of 79.19 Mb repetitive sequences, accounting for 42.73% of the genome. Total interspersed repeats (including 21.08% unclassified transposons) accounted for 39.27% of the whole genome, while tandem repeats represented 1.15% (Table 4). And detailed statistics of interspersed repeats reveal that DNA transposons being particularly abundant except the unclassified repeats, they are important for genome evolution, causing chromosomal recombination and genetic innovation[52]. The mean repeat coverage of each chromosome was calculated using 200 kbp sliding windows and is shown in Fig. 2A, along with the GC content.

Gene prediction and function annotation. After masking repeats in the genome, we performed gene prediction using a strategy that combined transcriptome-based, homology-based, and *de novo* prediction methods. For transcriptome-based prediction, libraries from female and male adults were constructed, and sequencing yielded 64.8 Gb of filtered RNA-seq data (Table 1). These validated data were aligned to the *C. ruficrus* genome using Hisat2 v2.1.0[53] with the parameters "–sensitive–no-mixed–no-discordant -X 1000 -I 1" to identify candidate gene regions, followed by transcript assembly using StringTie v1.0.4[54]. Full gene structure annotations and *de novo* gene prediction were carried out using BRAKER3[55,56]. For homology-based retrieval, Miniprot v2.2.26[57] was used to align protein sequences of clade-partitioned file (Arthropoda.fa) from OrthoDB v11[58] to the *C. ruficrus* genome. The results from the homology-based (weight score: 5), transcriptome-based (weight score: 10), and *de novo* predictions (weight score: 10) were integrated using EvidenceModeler v5.0.2[59] to generate a weighted and non-redundant consensus official gene set (OGS)[60].

Repeat Type	Number of elements	Length occupied (bp)	% of genome
Major interspersed repeats			
SINEs	4,403	547,913	0.3
LINEs	45,331	8,532,108	4.6
LTRs	32,996	8,167,443	4.41
DNA transposons	99,923	16,052,930	8.66
Rolling-circles	22,692	4,596,921	2.48
Unclassified	219,468	39,072,133	21.08
Major tandem repeats			
Small RNA	6,308	1,617,038	0.87
Satellites	2,575	278,530	0.15
Simple repeats	3,431	239,477	0.13

Table 4. Statistics of repetitive sequence of *Cotesia ruficrus* genome. Notes: most repeats fragmented by insertions or deletions have been counted as one element; SINEs, short interspersed nuclear elements; LINEs, long interspersed nuclear elements; LTR, long terminal repeat.

		Number of genes	Percentage
All protein-coding genes		14,001	100%
Annotated using databases	UniProt	8,790	62.78%
	GO	7,364	52.60%
	KEGG	3,998	28.56%
	Pfam	9,299	66.42%
	COG	9,399	67.13%

Table 5. Statistics of functionally annotated genes of *Cotesia ruficrus* genome.

A total of 14,007 genes were successfully predicted in the *C. ruficrus* genome, with an average exon number of 4.5. The average lengths of genes, coding sequences (CDS) and introns were 5,195 bp, 328 bp and 1,039 bp, respectively. Gene density on each chromosome (sliding window size: 20kbp) is showed in Fig. 2A. Furthermore, protein-coding genes were annotated using PANNZER2 database[61], and 62.78% of the genes were assigned to the Uniprot database. We also used the eggNOG-mapper web server[62] to identify and classify proteins and domains. As a result, 7,364 genes were assigned to GO terms, and 3,998 genes were mapped to at least one KEGG pathway (Table 5).

Data Records

The raw sequence data reported in this paper have been first deposited in the Genome Sequence Archive (GSA)[63] in National Genomics Data Center (NGDC)[64], China National Center for Bioinformation (CNCB)/Beijing Institute of Genomics, Chinese Academy of Sciences (BioProject accession number: PRJCA032354; GSA: CRA020489)[65] that are publicly accessible at https://ngdc.cncb.ac.cn/gsa.

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JBLUSQ000000000.1[66]. The version described in this paper is version JBLUSQ000000000.1. The genome assembly file has also been deposited in NGDC under accession number GWHFPUF00000000.1[39].

Technical Validation

Nucleic acid quality. The integrity of DNA samples was assessed by electrophoresis on 0.75% agarose gels to confirm the absence of degradation and contamination. Subsequently, the purity of the extracted DNA was evaluated using a NanoDrop™ One UV-V is spectrophotometer (Thermo Fisher Scientific, USA), with acceptable OD260/280 ratios falling within the range of 1.8 to 2.0 and OD260/230 ratios between 2.0 and 2.2. The precise concentration of DNA was determined using a Qubit® 3.0 Fluorometer (Life Technologies, CA, USA). For RNA, degradation and contamination of it were examined via 1% agarose gel electrophoresis. RNA concentration was quantified using the Qubit® RNA Assay Kit in conjunction with a Qubit Fluorometer. Additionally, RNA integrity was assessed using the RNA Nano 6000 Assay Kit on the Bioanalyzer 2100 system (Agilent Technologies, CA, USA).

Evaluation of genome assembly and annotation. To assess the quality of the genome assembly and annotation, we employed an evaluation from two aspects. For the genome assembly accuracy, PacBio long reads were aligned to the assembled genome using minimap2[67] with the parameter "-x map-pb". The mapping efficiency and genome coverage were subsequently evaluated using SAMtools v1.4[68]. The results indicated a high mapping rate of 99.99% for the long reads, achieving a mean genome coverage of 69.19X. Additionally, RNA-seq data from adult females and males were mapped to the genome of *C. ruficrus* using BWA-MEM2 v2.2.1[69], yielding a mapping rate of approximately 91.87%.

The completeness of the genome was evaluated from two distinct aspects: the assembled genomic sequences and the annotated protein sequences (Fig. 2A). The Benchmarking Universal Single-Copy Orthologs (BUSCO) v4.0.5[70] tool was utilized for this assessment, with the parameters "-l insecta_odb10 -g genome" (Fig. 2B).

Code availability

No custom code was used for this study. All data analyses were conducted using published bioinformatics software with default settings, unless otherwise noted.

Received: 16 December 2024; Accepted: 14 March 2025;
Published online: 26 March 2025

References

1. Zhang, Z. Q. A study on biological and ecological characteristics of *Apanteles ruficrus* Haliday. *Nat. Enemies Insects* 84–89 (1986).
2. Hill, M. G. Analysis of the biological control of *Mythimna separata* (Lepidoptera: Noctuidae) by *Apanteles ruficrus* (Braconidae: Hymenoptera) in New Zealand. *J. Appl. Ecol.* **25**, 197 (1988).
3. McCutcheon, G. S., Salley, W. Z. & Turnipseed, S. G. Biology of *Apanteles ruficrus*, an imported parasitoid of *Pseudoplusia includens*, *Trichoplusia ni*, and *Spodoptera frugiperda* (Lepidoptera: Noctuidae). *Environ. Entomol.* **12**, 1055–1058 (1983).
4. Chen, Y. *et al*. *Cotesia ruficrus* (Hymenoptera: Braconidae) parasitizing *Cnaphalocrocis medinalis* (Lepidoptera: Pyralidae): Developmental interactions and food utilization efficiency of hosts. *J. Econ. Entomol.* **109**, 588–593 (2016).
5. Gupta, A. *et al*. *Cotesia ruficrus* (Haliday, 1834) (Hymenoptera: Braconidae) emerging as a common natural parasitoid of *Spodoptera frugiperda* (J. E. Smith) (Lepidoptera: Noctuidae) in Indian maize fields. *J. Biol. Control* **33**, 193–196 (2019).
6. Hill, M. G. Effects of *Cotesia ruficrus* (Braconidae: Hymenoptera) parasitism and rearing density on *Mythimna separata* (Noctuidae: Lepidoptera) food consumption, and implications for biological control. *New Zealand J. Agric. Res.* **29**, 281–287 (1986).
7. Yu, D. S., van Achterberg, C. & Horstmann, K. Taxapad 2016 – World Ichneumonoidea 2011. *Taxonomy, Biology, Morphology and Distribution*. https://www.bioinfo.org.uk/html/b162939.htm (2016).
8. Goergen, G., Kumar, P. L., Sankung, S. B., Togola, A. & Tamò, M. First report of outbreaks of the fall armyworm *Spodoptera frugiperda* (J E Smith) (Lepidoptera, Noctuidae), a new alien invasive pest in west and central Africa. *PLOS ONE* **11**, e0165632 (2016).
9. Jacobs, A., Vuuren, Avan & Rong, I. H. Characterisation of the fall armyworm (*Spodoptera frugiperda* J.E. Smith) (Lepidoptera: Noctuidae) from South Africa. *Afr. Entomol.* **26**, 45–49 (2018).
10. Sun, X. *et al*. Case study on the first immigration of fall armyworm, *Spodoptera frugiperda* invading into China. *J. Integr. Agric.* **20**, 664–672 (2021).
11. Montezano, D. G. *et al*. Host plants of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) in the Americas. *Afr. Entomol.* **26**, 286–300 (2018).
12. Gurr, G. M. *et al*. Parasitoids of the rice leaffolder *Cnaphalocrocis medinalis* and prospects for enhancing biological control with nectar plants. *Agric. For. Entomol.* **14**, 1–12 (2012).
13. Li, X., An, J.-L., Wang, W.-J., Liu, T.-X. & Zhang, S.-Z. Influence of four noctuid pests on preference and offspring fitness of *Cotesia ruficrus*. *Crop Prot.* **188**, 107027 (2025).
14. Patil, S. S., Kamble, C. & Sathe, T. V. Biocontrol potential of *Cotesia ruficrus* Hal. (Hymenoptera: Braconidae) against different Lepidopterous pests. *Biolife* **4**, 343–346 (2022).
15. Pal, S. & Gupta, A. Severe outbreak of rice green semilooper, *Naranga aenescens* (Moore) (Lepidoptera: Noctuidae) along with its parasitoid complex in Sub-Himalayan West Bengal, India. *Entomol. News* **127**, 286–291 (2017).
16. Watanabe, C. Notes on braconid parasites of *Naranga aenescens* Moore occurring in Japan (Hymenoptera, Braconidae). *Insecta matsumurana* **28**, 131–132 (1966).
17. Zhu, P. *et al*. Selective enhancement of parasitoids of rice Lepidoptera pests by sesame (*Sesamum indicum*) flowers. *BioControl* **60**, 157–167 (2015).
18. Lou, Y.-G., Zhang, G.-R., Zhang, W.-Q., Hu, Y. & Zhang, J. Biological control of rice insect pests in China. *Biol. Control* **67**, 8–20 (2013).
19. Wang, W.-W., Han, K.-R., Jing, X.-F., Liu, T.-X. & Zhang, S.-Z. Odorant-binding protein CrufOBP1 in *Cotesia ruficrus* females plays a pivotal role in the detection of *Spodoptera frugiperda* larvae. *Int. J. Biol. Macromol.* **274**, 133491 (2024).
20. Han, K.-R. *et al*. Characterization of CrufCSP1 and its potential involvement in host location by *Cotesia ruficrus* (Hymenoptera: Braconidae), an indigenous parasitoid of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) in China. *Insects* **14**, 920 (2023).
21. Li, X. *et al*. Effect of *Cotesia ruficrus* parasitization on diversity and community composition of intestinal bacteria in *Spodoptera frugiperda*. *Insects* **15**, 570 (2024).
22. Yu, W. *et al*. Comprehensive assessment of 11 *de novo* HiFi assemblers on complex eukaryotic genomes and metagenomes. *Genome Res.* **34**, 326–340 (2024).
23. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).
24. Whibley, A., Kelley, J. L. & Narum, S. R. The changing face of genome assemblies: Guidance on achieving high-quality reference genomes. *Mol. Ecol. Resour.* **21**, 641–652 (2021).
25. Gauthier, J. *et al*. Chromosomal scale assembly of parasitic wasp genome reveals symbiotic virus colonization. *Commun. Biol.* **4**, 104 (2021).
26. Ye, X. *et al*. Genome of the parasitoid wasp *Cotesia chilonis* sheds light on amino acid resource exploitation. *BMC Biol.* **20**, 118 (2022).
27. Ye, X., Yang, Y., Zhao, X., Fang, Q. & Ye, G. The state of parasitoid wasp genomics. *Trends Parasitol.* **40**, 914–929 (2024).
28. Ye, X., Yang, Y., Fang, Q. & Ye, G. Genomics of insect natural enemies in agroecosystems. *Curr. Opin. Insect. Sci.* **68**, 101298 (2024).
29. Cruf_RNA4OGS_AM, *Genome Sequence Archive*, https://ngdc.cncb.ac.cn/gsa/search?searchTerm=CRX1264894 (2025).
30. Cruf_RNA4OGS_AF, *Genome Sequence Archive*, https://ngdc.cncb.ac.cn/gsa/search?searchTerm=CRX1264895 (2025).
31. Dobin, A. *et al*. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
32. Cruf_PacBioreads, *Genome Sequence Archive*, https://ngdc.cncb.ac.cn/gsa/search?searchTerm=CRX1264892 (2025).
33. Cruf_HiCreads, *Genome Sequence Archive*, https://ngdc.cncb.ac.cn/gsa/search?searchTerm=CRX1264893 (2025).
34. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
35. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**, 1432 (2020).
36. Hu, J. *et al*. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. *Genome Biol.* **25**, 107 (2024).
37. Chaisson, M. J. & Tesler, G. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinform.* **13**, 238 (2012).

38. Hu, J., Fan, J., Sun, Z. & Liu, S. NextPolish: A fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **36**, 2253–2255 (2020).
39. Zhao, X. Cruf_HiCgenome. Genome Warehouse. https://ngdc.cncb.ac.cn/gwh/Assembly/91414/show (2024).
40. Chen, S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *Imeta* **2**, e107 (2023).
41. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
42. Durand, N. C. *et al*. Juicer Provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
43. Zhang, Y., Xiong, Y. & Xiao, Y. 3dDNA: A computational method of building DNA 3D Structures. *Molecules* **27**, 5936 (2022).
44. Robinson, J. T. *et al*. Juicebox.js Provides a cloud-based visualization system for Hi-C Data. *Cell Syst.* **6**, 256–258.e1 (2018).
45. Belle, E. *et al*. Visualization of polydnavirus sequences in a parasitoid wasp chromosome. *J. Virol.* **76**, 5793–5796 (2002).
46. Gokhman, V. E. *Karyotypes of Parasitic Hymenoptera*. (Springer, Dordrecht, 2009).
47. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* Chapter 4, 4.10.1–4.10.14 (2009).
48. Flynn, J. M. *et al*. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. USA.* **117**, 9451–9457 (2020).
49. Bao, Z. & Eddy, S. R. Automated *de novo* identification of repeat sequence families in sequenced genomes. *Genome Res.* **12**, 1269–1276 (2002).
50. Price, A. L., Jones, N. C. & Pevzner, P. A. *De novo* identification of repeat families in large genomes. *Bioinformatics* **21**(Suppl 1), i351–358 (2005).
51. Camacho, C. *et al*. BLAST+: Architecture and applications. *BMC Bioinform.* **10**, 421 (2009).
52. Feschotte, C. & Pritham, E. J. DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41**, 331–68 (2007).
53. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
54. Pertea, M. *et al*. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
55. Hoff, K. J., Lomsadze, A., Borodovsky, M. & Stanke, M. Whole-genome annotation with BRAKER. *Methods Mol. Biol.* **1962**, 65–95 (2019).
56. Gabriel, L. *et al*. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. *bioRxiv* 2023.06.10.544449, https://doi.org/10.1101/2023.06.10.544449 (2024).
57. Li, H. Protein-to-genome alignment with miniprot. *Bioinformatics* **39**, btad014 (2023).
58. Kuznetsov, D. *et al*. OrthoDB v11: Annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Res.* **51**, D445–D451 (2023).
59. Haas, B. J. *et al*. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.* **9**, R7 (2008).
60. Zhao, X. Genome assembly and annotation of *Cotesia ruficrus*. *figshare. Dataset.* https://doi.org/10.6084/m9.figshare.27853896.v1 (2025).
61. Törönen, P., Medlar, A. & Holm, L. PANNZER2: A rapid functional annotation web server. *Nucleic Acids Res.* **46**, W84–W88 (2018).
62. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic Scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).
63. Chen, T. *et al*. The genome sequence archive family: toward explosive data growth and diverse data types. *Genom. Proteom. Bioinf.* **19**, 578–583 (2021).
64. CNCB-NGDC Members and Partners. Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2024. *Nucleic Acids Res.* **52**, D18–D32 (2024).
65. *National Genomics Data Center (NGDC) Genome Sequence Archive* https://ngdc.cncb.ac.cn/gsa/browse/CRA020489 (2024).
66. Zhao, X. *Cotesia ruficrus* isolate zhejiang, whole genome shotgun sequencing project. *GenBank.* https://www.ncbi.nlm.nih.gov/nuccore/JBLUSQ000000000.1 (2025).
67. Li, H. Minimap and miniasm: Fast mapping and *de novo* assembly for noisy long sequences. *Bioinformatics* **32**, 2103–2110 (2016).
68. Li, H. *et al*. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
69. Jung, Y. & Han, D. BWA-MEME: BWA-MEM emulated with a machine learning approach. *Bioinformatics* **38**, 2404–2413 (2022).
70. Seppey, M., Manni, M. & Zdobnov, E. M. BUSCO: Assessing genome assembly and annotation completeness. *Methods Mol. Biol.* **1962**, 227–245 (2019).

Acknowledgements

This work was supported by Program of National Natural Science Foundation of China (NSFC) (Grant No. 32202376 and Grant No. 32472631 to XHY, Grant No. 32302337 to XXZ), the China Postdoctoral Science Foundation (Grant No. 2022M722804 to XXZ), the Joint Funds of the National Natural Science Foundation of China (NSFC) (Grant No. U21A20225 to GYY), the Fundamental Research Funds for the Central Universities (226-2024-00070 to GYY).

Author contributions

Conceptualization and supervision: Gongyin Ye, Xinhai Ye; Sampling, experiments, and analysis: Xianxin Zhao, Yuanyuan Liu, Fang Wang and Qi Fang; Writing and editing: Xianxin Zhao, Xinhai Ye and Gongyin Ye.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to G.Y.

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.