



OPEN

DATA DESCRIPTOR

# Chromosome-level genome assembly of the flower thrips *Frankliniella intonsa*

Zhijun Zhang<sup>1</sup>✉, Jiandong Bao<sup>1</sup>, Qizhang Chen<sup>1,3</sup>, Jianyun He<sup>1</sup>, Xiaowei Li<sup>1</sup>, Jiahui Zhang<sup>1,2</sup>, Zhixing Liu<sup>1</sup>, Yixuan Wu<sup>1</sup>, Xuesheng Li<sup>3</sup>, Yunsheng Wang<sup>2</sup> & Yaobin Lu<sup>1</sup>✉

As an economically important insect pest, the flower thrips *Frankliniella intonsa* (Trybom) causes great damage to host plants by directly feeding and indirectly transmitting various pathogenic viruses. The lack of a well-assembled genomic resource has hindered our understanding of the genetic basis and evolution of *F. intonsa*. In this study, we used Oxford Nanopore Technology (ONT) long reads and High-throughput chromosome conformation capture (Hi-C) linked reads to construct a high-quality reference genome assembly of *F. intonsa*, with a total size of 225.5 Mb and a contig N50 of 3.37 Mb. By performing the Hi-C analysis, we anchored 91.68% of the contigs into 15 pseudochromosomes. Genomic annotation uncovered 17,581 protein-coding genes and identified 20.09% of the sequences as repeat elements. BUSCO analysis estimated over 98% of genome completeness. Our study is at the first time to report the chromosome-scale genome for the species of the genus *Frankliniella*. It provides a valuable genomic resource for further biological research and pest management of the thrips.

## Background & Summary

The flower thrips, *Frankliniella intonsa* (Trybom) (Thysanoptera: Thripidae), is a small-sized insect pest, well known for feeding and dwelling on the flower of host plants. It is widely distributed in the world including Europe, Asia, Oceania, and North America<sup>1,2</sup>, and becoming the dominant thrips species in several areas of China<sup>3,4</sup>. By rapid development and both sexual and parthenogenetic reproductive modes, *F. intonsa* is able to cause severe damage to various commercial crops, such as cowpea, eggplant<sup>5–9</sup>, and rapidly accumulated resistance to insecticides like spinosad<sup>10</sup>. In addition to direct damage to the leaves, flowers and fruits, *F. intonsa* is also capable of transmitting a variety of plant pathogenic viruses, such as Tomato spotted wilt orthotospovirus (TSWV) and Chrysanthemum stem necrosis virus (CSNV) to host plants<sup>11,12</sup>, resulting in destructive damage and huge economic losses every year. Interestingly, we found that the endosymbiont *Wolbachia* is dwelling in *F. intonsa*, but is absent in the sibling species *F. occidentalis* (unpublished data). High-quality genomic resources are urgently needed to elucidate the key genetic mechanisms of flower thrips like virus transmission, pheromone biosynthesis, insecticide resistance and bacterial mutualism. To date, despite the large number of species in the family Thripidae (true thrips), only five species have publicly available genomes, including a scaffold-level genome assembly of western flower thrips (*F. occidentalis*, 415.8 Mb)<sup>13</sup>, a chromosome-level genome assembly of melon thrips (*Thrips palmi*)<sup>14</sup>, a scaffold-level genome assembly of tobacco thrips (*F. fusca*, 370 Mb)<sup>15</sup>, and two chromosome-level genome assembly of bean blossom thrips (*Megalurothrips usitatus*) reported by Ma *et al.*<sup>16</sup> (238.14 Mb) and our group<sup>17</sup> (247.82 Mb), respectively. However, there is no reported genome assembly for *F. intonsa*, nor chromosome-level genome for the *Frankliniella* genus species. In this study, we report a high-quality chromosome-level genome assembly of *F. intonsa* using an integrated sequencing strategy including ONT, Illumina, and Hi-C. Our research provides valuable resources for studying the evolutionary genetics and molecular basis of *F. intonsa*.

To obtain a high-quality genome assembly of *F. intonsa*, a total of 31.63 Gb ONT long reads (~124-fold coverage) and 15.71 Gb NGS short reads (~62-fold coverage, 2 × 150 bp) were generated. Using the integrated

<sup>1</sup>State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-Products, Institute of Plant Protection and Microbiology, Zhejiang Academy of Agricultural Sciences, Hangzhou, 310021, China. <sup>2</sup>Hunan Provincial Key Laboratory for Biology and Control of Plant Diseases and Insect Pests, Hunan Agricultural University, Changsha, 410125, China. <sup>3</sup>Guangxi key laboratory of Agric-Environment and Agric-products Safety, Guangxi University, Nanning, Guangxi, 530004, China. ✉e-mail: [zhangzj@zaas.ac.cn](mailto:zhangzj@zaas.ac.cn); [lvyb@zaas.ac.cn](mailto:lvyb@zaas.ac.cn)

Genome features	<i>F. intonsa</i>
Total, size of contig assembly (Mb)	225.5
Number of contig	405
Contig N50 (bp)	3,373,233
Contig N90 (bp)	161,309
Number of pseudochromosomes	15
Scaffold N50 (bp)	13,489,000
Scaffold N90 (bp)	9,782,500
GC content	52.73%
Repeat content	20.09%
Number of protein coding genes	17,581
Average gene length (bp)	3,293
Average exon number Per gene	6.28

**Table 1.** Statistics of *F. intonsa* reference genome.

sequencing data, we obtained a contig-level genome with a total size of 225.5 Mb, consisting of 405 contigs with N50 length of 3.37 Mb and N90 length of 161 Kb. (Table 1). The total length of the genome assembly is similar to the estimated genome size (approximately 234.5 Mb) based on 21-mer depth analysis (Fig. 1a). The total GC content is 52.73%, which is comparable to the other Thripidae species<sup>13–17</sup>. To improve the continuity of the genome assembly, we exploited 41.97 Gb (~165-fold coverage) Hi-C data, which generated about 57 million Hi-C contacts, to concatenate the contigs. Approximately 91.68% of the contig sequence was successfully anchored into 15 pseudochromosomes ranging from 9.78 Mb to 20.82 Mb (Fig. 1b,c). We further performed BUSCO analysis to assess the completeness of the genome assembly based on four categories of datasets, including Eukaryota, Metazoa, Arthropoda and Insecta (odb\_10). As a result, 96.9% conserved Eukaryotic genes and more than 98% of the core genes in the other three datasets were identified, strongly suggesting a high level of completeness of the *F. intonsa* genome assembly (Fig. 1d).

Using multiple repeat annotation software, we constructed a repeat library containing 1,347 repeat consensus sequences in the *F. intonsa* genome. We then performed a genome-wide scan of repeat-associated regions based on the repeat library. As a result, we annotated approximately 20.09% repeat regions in the *F. intonsa* genome (Table 1). Remarkably, the repeat content in the *F. intonsa* genome is significantly higher than in the other published genomes of Thripidae species (*F. occidentalis* (9.86%), *M. usitatus* (15.05%), *T. palmi* (6.45%)), suggesting the amplification of repeat elements in the *F. intonsa* genome<sup>13–17</sup>.

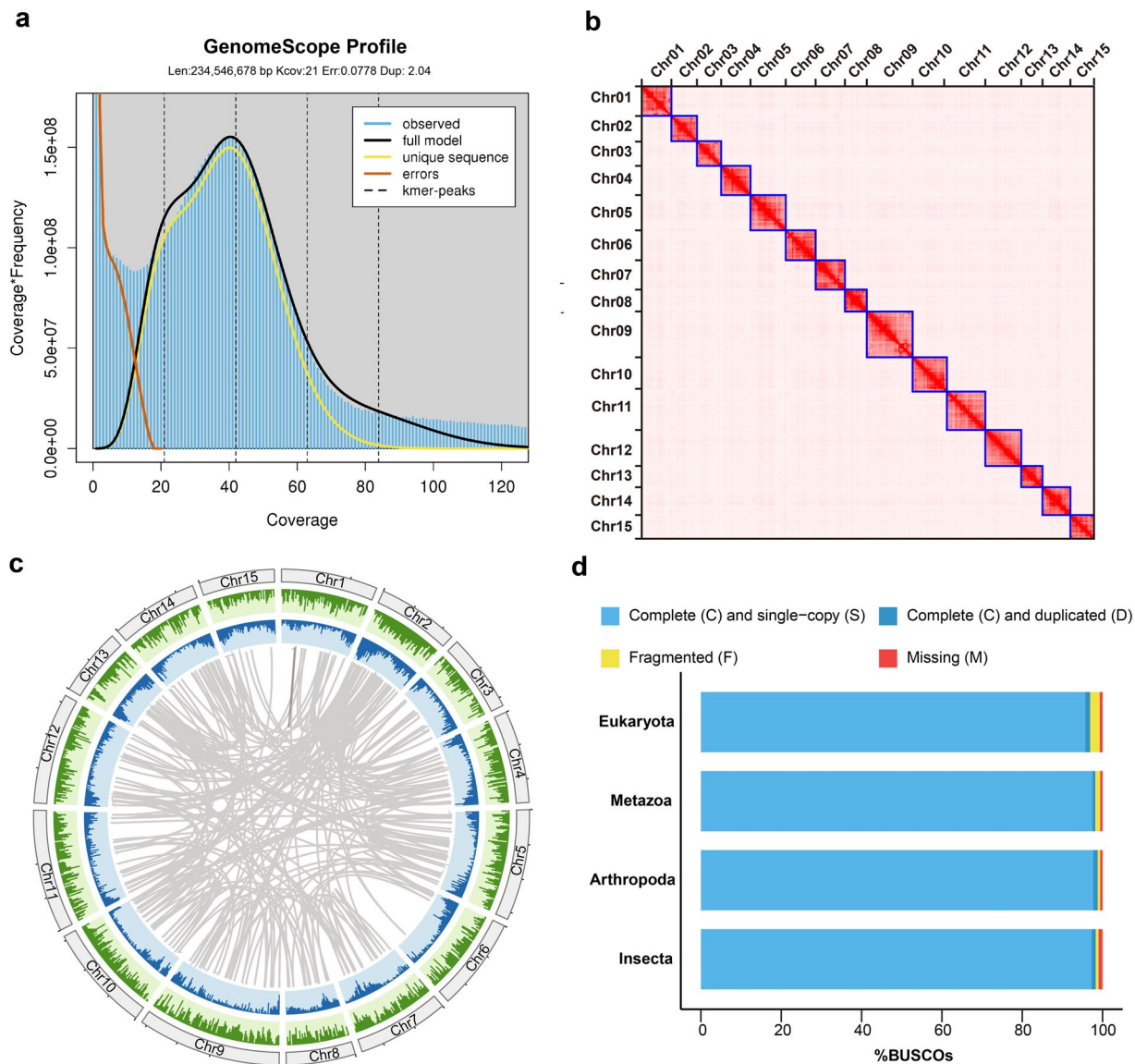
A combined approach of ab initio prediction, homolog-based prediction, and transcript-based prediction was used to predict gene structure in the *F. intonsa* genome. This resulted in a total of 17,581 protein-coding genes, which is comparable to other Thripidae species. BUSCO analysis using the gene model showed that 89.5%, 93.8%, 94.4% and 94.0% of the core genes from the Eukaryota, Metazoa, Arthropoda and Insecta datasets, respectively, were complete, indicating a high level of completeness and credibility of the gene prediction results (Fig. 2). Then, we functionally annotated the protein-coding genes based on five major databases, including Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Pfam, Clusters of Orthologous Genes (COG) and the Carbohydrate-Active enZYmes (CAZy). More than half of the genes (58.42%, 10,272/17,581) were well annotated with at least one functional result (Fig. 3).

## Methods

**Sampling, extraction, and sequencing.** *F. intonsa* populations were collected on pepper (*Capsicum annuum* L.) from Jiaying (30.75° N, 120.79° E), Zhejiang, China, in 2017, and reared on fresh bean, *Phaseolus vulgaris*, in a climate-controlled chamber (25 ± 1 °C, 16 L). Approximately 200 adult *F. intonsa* samples from the laboratory population with mixed ages were decontaminated by immersion in 1% sodium hypochlorite solution (Gaide chemical, Hangzhou, Zhejiang, China) for 5 min, and followed by rinsing with sterile water and then immersion in 70% ethanol and finally rinsing with sterile water again. Samples were snap frozen in liquid nitrogen and stored at –80 °C.

Genomic DNA and mRNA were extracted and purified using QIAGEN DNA/RNA tissue kit (QIAGEN 69506/73404, Hilden, Germany), and prepared for sequencing libraries according to the manufacturer's instructions for sequencing technology (Nextomics Biosciences Co., Ltd, Wuhan, China). Long DNA fragments were sequenced on the Oxford Nanopore PromethION platform, and short-read sequencing, Hi-C sequencing and RNA-seq were performed on the Illumina NovaSeq. 6000 platform (Table S1).

**Genome assembly, scaffolding, completeness evaluation.** The raw Oxford nanopore sequencing technology (ONT) reads were quality controlled using LongQC (v1.2.0c)<sup>18</sup> to remove low-quality, barcode and adapter sequences. The clean ONT reads were then used for *de novo* assembly using NextDenovo (v2.5.2)<sup>19</sup> with an expected genome size of 250 Mb, and the draft contig assembly was polished by NextPolish (v1.4.1)<sup>20</sup> with default parameters using both of the clean ONT reads and NGS short reads. To obtain chromosome-level data, we exploited Hi-C sequencing data to concatenate the contigs based on chromosomal interaction signals. Bowtie2 (v2.5.1)<sup>21</sup> and 3D-DNA (v180114)<sup>22</sup> were used for scaffolding, and Juicebox (v2.20.00)<sup>23</sup> was then used for manual scaffold correction. To assess the completeness of genome assembly, we performed BUSCO (v5.4.4)<sup>24</sup> analysis

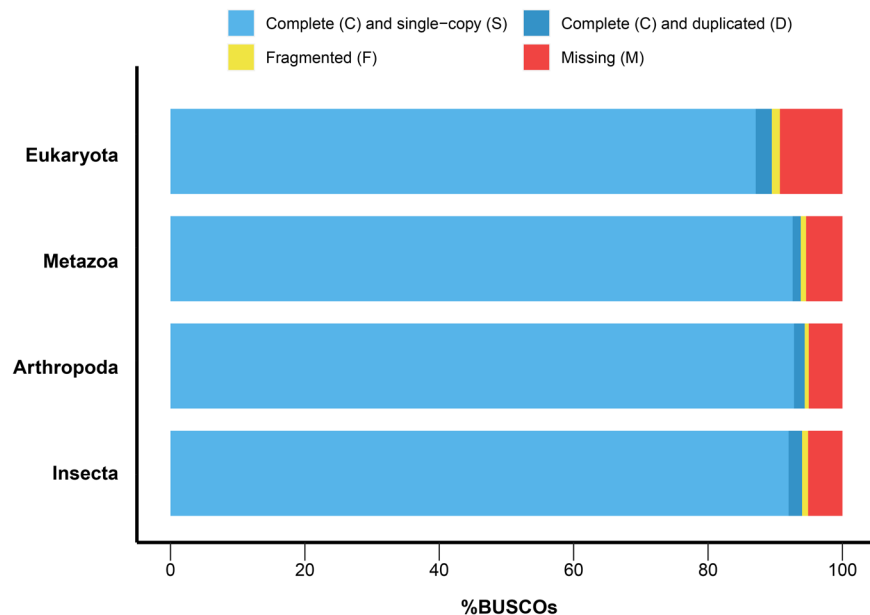


**Fig. 1** Characteristics of the *F. intonsa* genome. **(a)** Estimation of genome size based on NGS reads using 21-kmer analysis. **(b)** The heatmap represents 15 pseudochromosomes of the *F. intonsa* genome. **(c)** The circo plot describes the genomic characteristics of *F. intonsa*, including the chromosomes (the outer circle), gene density (green cycle), repeat density (blue circle), and paralogous genes (line plot). **(d)** BUSCO analysis based on the Eukaryota, Metazoa, Arthropoda and Insecta (odb\_10) data sets.

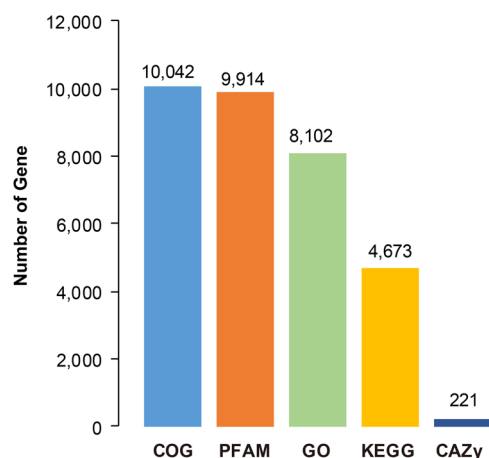
with the genome model by using the Eukaryota, Metazoa, Arthropoda and Insecta (odb10) datasets (<https://busco-data.ezlab.org/v5/data/lineages/>). Clean NGS and ONT reads were realigned to the genome assembly using bwa-mem2 (v2.2.1)<sup>25</sup> and minimap2 (v2.26)<sup>26</sup>, respectively.

**Repeat annotation.** Repeatmodeler (v2.0.3)<sup>27</sup> in combination with LTR\_finder (v1.07)<sup>28</sup>, LTRharvest (v1.6.2)<sup>29</sup> and LTR\_retriever (v2.9.0)<sup>30</sup> were used to identify repeat elements, including LTR retrotransposons, non-LTR retrotransposons, DNA transposons and microsatellites. The *de novo* repeat library was then used as a seed for RepeatMasker (v4.1.0)<sup>31</sup> to find and mask all repeat regions in the final genome assembly.

**Gene structural and functional annotation.** We exploited several software programs including Funannotate (v1.5.3)<sup>32</sup>, Fgenesh<sup>33</sup>, Exonerate (v2.2.0)<sup>34</sup>, and Trinity (v2.11.0)<sup>35</sup> to predict gene structure. Finally, EvidenceModeler (v2.0.0)<sup>36</sup> was used to combine *de novo*, homolog-based and transcriptome-based gene predictions into a final gene structure dataset. Two software, eggNOG-mapper (v2.1.9)<sup>37</sup> and Interproscan (v5.62-94.0)<sup>38</sup>, in combination with five databases including KOG, KEGG, CAZys, Pfam and Gene Ontology (GO) were used to functionally annotate the genes.



**Fig. 2** BUSCO analysis of the gene models based on the Eukaryota, Metazoa, Arthropoda and Insecta (odb\_10) data sets.



**Fig. 3** The number of genes annotated with COG, PFAM, GO, KEGG and CAZy databases.

### Data Records

The genome sequence and annotation data were deposited in the Genome Warehouse (GWH, <https://ngdc.cnbc.ac.cn/gwh>) at the National Genomic Data Center (NGDC)<sup>39</sup>, under the accession number of GWHDOOB00000000. Raw data from Nanopore (CRR824223<sup>40</sup>), Illumina (CRR825014<sup>41</sup>, CRR825015<sup>42</sup>) and Hi-C (CRR824225<sup>43</sup>) genome sequencing and RNA-seq (CRR824226<sup>44</sup>) were deposited in the Genome Sequence Archive<sup>45</sup> (GSA, <https://ngdc.cnbc.ac.cn/gsa>) at the NGDC. All data were related to the BioProject PRJCA018338. The genome sequence and raw reads were also deposited at WGS (JAWJED000000000.1<sup>46</sup>) and GSA (SRR26384730<sup>47</sup> for Nanopore, SRR26384729<sup>48</sup> for Illumina, SRR26384728<sup>49</sup> for Hi-C, and SRR26384727<sup>50</sup> for RNA-seq data) at NCBI, respectively, under BioProject PRJNA1027977.

### Technical Validation

Two different strategies were used to evaluate the completeness and accuracy of the *F. intonsa* genome. First, BUSCO analysis based on the Eukaryota, Metazoa, Arthropoda and Insecta (odb\_10) datasets revealed that 96.9%, 98.2%, 98.8% and 98.3% of the core genes were successfully identified as complete. Second, we re-aligned the NGS, ONT and RNA-seq reads to the *F. intonsa* genome with the mapping rates of 92.80%, 90.48% and 88.63%, respectively. For evaluation of gene prediction completeness and accuracy, we performed BUSCO analysis based on the Eukaryota, Metazoa, Arthropoda and Insecta (odb\_10) datasets, which resulted in 89.5%, 93.8%, 94.4% and 94.0% of completeness, respectively.

## Code availability

No specific script was used in this work. The codes and pipelines used for data processing were all executed according to the manual and protocols of the corresponding bioinformatics softwares (detail parameters see Table S2).

Received: 20 July 2023; Accepted: 20 November 2023;

Published online: 30 November 2023

## References

- Ullah, M. S. & Lim, U. T. Life History Characteristics of *Frankliniella occidentalis* and *Frankliniella intonsa* (Thysanoptera: Thripidae) in Constant and Fluctuating Temperatures. *J. Econ. Entomol.* **108**, 1000–1009 (2015).
- Jones, D. R. Plant Viruses Transmitted by Thrips. *Eur. J. Plant Pathol.* **113**, 119–157 (2005).
- Mao, L. *et al.* Attraction effect of different colored cards on thrips *Frankliniella intonsa* in cowpea greenhouses in China. *Sci. Rep.* **8**, 13603 (2018).
- Zhang, P., Zhu, X. & Lu, Y. Behavioural and chemical evidence of a male-produced aggregation pheromone in the flower thrips *Frankliniella intonsa*. *Physiol. Entomol.* **36**, 317–320 (2011).
- Lewis T. *Thrips, their biology, ecology and economic importance* (London Academic Press. 1973).
- Wei, S., Lu, D., Qu, Y. & Zhang, Q. Efficacy trials of five pesticides against thrips from mango and cowpea. *J. Environ. Entomol.* **34**, 519–524 (2012).
- Akella, S. V. S. *et al.* Identification of the Aggregation Pheromone of the Melon Thrips, *Thrips palmi*. *PLoS One* **9**, e103315 (2014).
- Wang, C., Lin, F., Chiu, Y. & Shih, H. Species of *Frankliniella Trybom* (Thysanoptera: Thripidae) from the Asian-Pacific Area. *Zool. Stud.* **49**, 824–848 (2010).
- Lim, U. T., Kim, E. & Mainali, B. P. Flower model traps reduced thrips infestations on a pepper crop in field. *J. Asia-Pac. Entomol.* **16**, 143–145 (2013).
- Hiruta, E., Aizawa, M., Nakano, A. & Sonoda, S. Nicotinic acetylcholine receptor  $\alpha 6$  subunit mutation (G275V) found in a spinosad-resistant strain of the flower thrips, *Frankliniella intonsa* (Thysanoptera: Thripidae). *J. Pestic. Sci.* **43**, 272–276 (2018).
- Prins, M. & Goldbach, R. The emerging problem of tospovirus infection and nonconventional methods of control. *Trends Microbiol.* **6**, 31–35 (1998).
- Riley, D. G., Joseph, S. V., Srinivasan, R. & Diffie, S. Thrips Vectors of Tospoviruses. *J. Integr. Pest Manag.* **2**, 11–110 (2011).
- Rotenberg, D. *et al.* Genome-enabled insights into the biology of thrips as crop pests. *BMC Biol.* **18**, 142 (2020).
- Guo, S.-K. *et al.* Chromosome-level assembly of the melon thrips genome yields insights into evolution of a sap-sucking lifestyle and pesticide resistance. *Mol. Ecol. Resour.* **20**, 1110–1125 (2020).
- Catto, M. A. *et al.* Pest status, molecular evolution, and epigenetic factors derived from the genome assembly of *Frankliniella fusca*, a thysanopteran phyto virus vector. *BMC Genomics* **24**, 343 (2023).
- Ma, L. *et al.* Chromosome-level genome assembly of bean flower thrips *Megalurothrips usitatus* (Thysanoptera: Thripidae). *Sci. Data* **10**, 252 (2023).
- Zhang, Z. *et al.* The chromosome-level genome assembly of Bean blossom thrips (*Megalurothrips usitatus*) reveals an expansion of protein digestion-related genes in adaption to high-protein host plants. *Int. J. Mol. Sci.* **24**, 11268 (2023).
- Fukawasa, Y., Ermini, L., Wang, H., Carty, K. & Cheung, M. S. LongQC: A Quality Control Tool for Third Generation Sequencing Long Read Data. *G3-Genes Genomes Genet.* **10**, 1193–1196 (2020).
- Hu J, *et al.* An efficient error correction and accurate assembly tool for noisy long reads. Preprint at <https://www.biorxiv.org/content/10.1101/2023.03.09.531669v1.full> (2023).
- 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).
- Giannoulatou, E., Park, S. H., Humphreys, D. T. & Ho, J. W. Verification and validation of bioinformatics software without a gold standard: a case study of BWA and Bowtie. *BMC Bioinformatics* **15**, S15 (2014).
- Dudchenko, O. *et al.* De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
- Robinson, J. T. *et al.* Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data. *Cell Syst.* **6**, 256–258.e1 (2018).
- Seppy, M., Manni, M. & Zdobnov, E. M. BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods in molecular biology* **1962**, 227–245 (2019).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
- Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. USA* **117**, 9451–9457 (2020).
- Xu, Z. & Wang, H. LTR\_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).
- Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**, 18 (2008).
- Ou, S. & Jiang, N. LTR\_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons. *Plant Physiol.* **176**, 1410–1422 (2018).
- Chen, N. Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. *Curr Protoc Bioinformatics.* **5**, 4.10.1–4.10.14 (2004).
- Palmer, J. & Stajich, J. Switzerland nextgenusfs/funcannotate: funcannotate v1.5.3 (Version 1.5.3). *Zenodo*. <https://doi.org/10.5281/zenodo.2604804> (2019).
- Solovyev, V., Kosarev, P., Seledsov, I. & Vorobyev, D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol.* **7**, S10 (2006).
- Slater, G. S. C. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005).
- Grabherr, M. G. *et al.* Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* **29**, 644 (2011).
- Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
- Huerta-Cepas, J. *et al.* Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).
- Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
- Chen, M. *et al.* Genome Warehouse: a public repository housing genome-scale data. *Genomics, proteomics & bioinformatics* **19**, 584–589 (2021).
- NGDC Genome Sequence Archive (GSA) <https://ngdc.cnpc.ac.cn/gsa/browse/CRA011845/CRR824223> (2023).
- NGDC Genome Sequence Archive (GSA) <https://ngdc.cnpc.ac.cn/gsa/browse/CRA011862/CRR825014> (2023).

42. NGDC Genome Sequence Archive (GSA) <https://ngdc.cnbc.ac.cn/gsa/browse/CRA011862/CRR825015> (2023).
43. NGDC Genome Sequence Archive (GSA) <https://ngdc.cnbc.ac.cn/gsa/browse/CRA011845/CRR824225> (2023).
44. NGDC Genome Sequence Archive (GSA) <https://ngdc.cnbc.ac.cn/gsa/browse/CRA011845/CRR824226> (2023).
45. Chen, T. *et al.* The Genome Sequence Archive Family: Toward Explosive Data Growth and Diverse Data Types. *Genomics Proteomics & Bioinformatics*. **19**, 578–583 (2021).
46. Zhang, Z. J. *Frankliniella intonsaisolate* FiZJ1, whole genome shotgun sequencing project. GenBank <https://identifiers.org/ncbi/insdc:JAWJED000000000.1> (2023).
47. NCBI Sequence Read Archive, <https://identifiers.org/ncbi/insdc.sra:SRR26384730> (2023).
48. NCBI Sequence Read Archive, <https://identifiers.org/ncbi/insdc.sra:SRR26384729> (2023).
49. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR26384728> (2023).
50. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR26384727> (2023).

## Acknowledgements

This work was supported by the National Natural Science Foundation of China (31672031 and 32272537); the Key R&D Program of China (2022YFD1401204, 2022YFC2601405); the Key Research and Development Program of Zhejiang Province, China (2021C02003) and the Collaborative Promotion Program of Agricultural Major Technology(2023ZDXT04-3).

## Author contributions

Z.Z. conceived the idea, analyzed the data, wrote the original draft, revised the manuscript, and got the funding; J.B., Y.W. (Yunsheng Wang), and J.Z. analyzed the data; Q.C., J.H., Z.L., X.L. (Xiaowei Li), X.L. (Xuesheng Li) and Y.W. (Yixuan Wu) prepared the materials; Y.L. supervised the project; All authors have read and agreed to the published version of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41597-023-02770-3>.

**Correspondence** and requests for materials should be addressed to Z.Z. or Y.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023