scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Chromosome-level genome assembly from a single green slender planthopper *Saccharosydne procerus*

Xin-Yang Liu[1,2,4], Zhuo-Qi Liu[2,4], Xun Zhou[3,4], Ni-Tong Xu[2], Jie Li[2], Hai-Jun Xu[2] ✉ & Jue-Feng Zhang[1] ✉

The green slender planthopper (*Saccharosydne procerus*) is a significant pest for water bamboo (*Zizania latifolia*), which is an economically important aquatic vegetable cultivated in East Asia for its edible swollen shoot. Despite its agricultural impact, the molecular basis of *S. procerus*'s ecological adaptations remains poorly understood. To address this gap, we assembled the first reference genome of *S. procerus* at the chromosomal level using ultra-low input PacBio HiFi libraries and Hi-C sequencing. The genome size is 752.84 Mb, with a contig N50 size of 1.88 Mb and scaffold N50 size of 63.55 Mb. A total of 97.52% of assembled sequences were anchored to 14 pseudo-chromosomes, and BUSCO analysis yielded an Insecta completeness score of 94.8%. A total of 24,349 protein-coding genes were annotated and 310.78 Mb repetitive sequences occupying 41.28% of genome were pinpointed. Our study presents the first high-quality genome assembly of *S. procerus*, laying the foundation for further studies on its genetic characteristics and pest management.

Background & Summary

Water bamboo *Zizania latifolia* (Griseb.) belonging to the Oryzeae of the Gramineae, is an asexual aquatic vegetable with an edible swelling gall (Fig. 1a). Water bamboo is cultivated at scattered locations in some east Asian countries such as Japan, South Korea, and China for more than 2000 years[1,2]. The green slender planthopper *Saccharosydne procerus* (Hemiptera: Delphacidae) is a major pest for water bamboo. Both *S. procerus* adults and nymphs feed on sap and thus cause leaf chlorosis and stunted growth, resulting in yield losses of up to 80%[3]. This pest also infests rice but shows strong preference for water bamboo in a rice-water bamboo co-cultivated system[4]. Additionally, honeydew secreted by green slender planthoppers facilitates sooty mold proliferation and increase host susceptibility to secondary pathogens, including *Helminthosporium* leaf spot and rust fungi[5]. Current management strategies for green slender planthoppers predominantly rely on chemical pesticides, but the efficacy of chemical control is significantly diminished because of its small body size (3–7 mm) (Fig. 1a), high mobility, and rapid reproductive rate[6]. In addition, extensive application of pesticides poses risks to human health, the environment, and pesticide resistance evolution.

Although *S. procerus* is an insect pest for water bamboo, it plays a key ecological role in rice-water bamboo co-cultivated paddies: its eggs serve as an alternative host for the parasitic wasp *Anagrus nilaparvatae* (Hymenoptera: Mymaridae). In rice paddies, *A. nilaparvatae* is an important biological control agent for the brown planthopper *Nilaparvata lugens*, the most destructive insect pest of rice in China[7,8]. The rice-water bamboo co-cultivation system helps stabilize wasp populations when rice fields are disturbed by harvesting or pesticide use[4]. However, a lack of genomic resources for *S. procerus* has hindered our understanding of its population genetics, adaptive mechanisms, and interactions with natural enemies. This knowledge gap has limited the development of effective, sustainable pest management strategies.

[1]Institute of Plant Protection and Microbiology, Zhejiang Academy of Agricultural Sciences, Hangzhou, 310021, China. [2]State Key Laboratory of Rice Biology and Breeding, Institute of Insect Sciences, Zhejiang University, Hangzhou, 310058, China. [3]Annoroad Gene Technology (Beijing) Co., Ltd, Beijing, 100180, China. [4]These authors contributed equally: Xin-Yang Liu, Zhuo-Qi Liu, Xun Zhou. ✉e-mail: haijunxu@zju.edu.cn; zhangjf@zaas.ac.cn

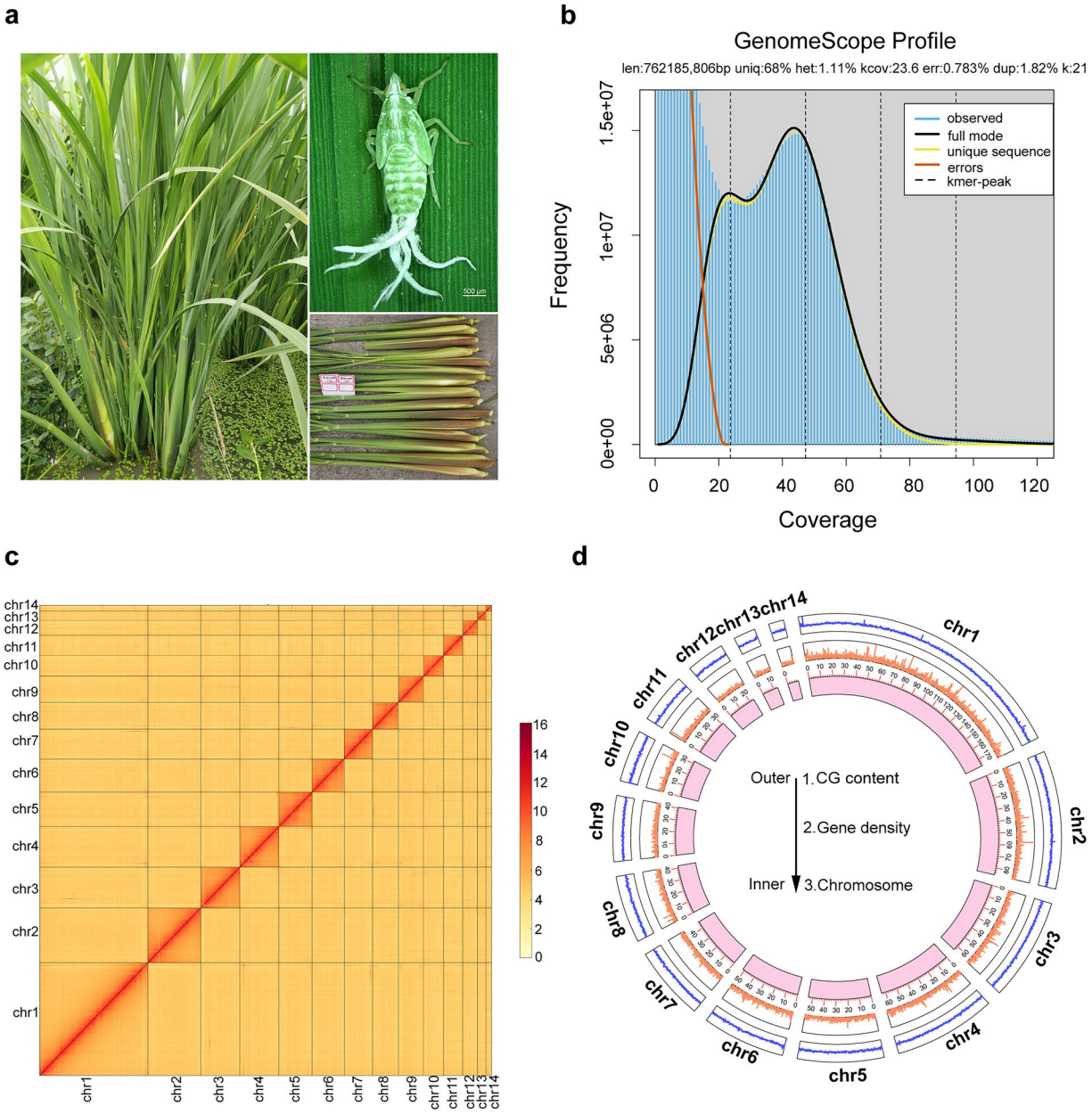

Fig. 1 Characteristics of the *S. procerus* genome. (**a**) water bamboo (*Z. latifolia*) with a swelling gall and the slender planthopper S. *procerus*. (**b**) *S. procerus* genome size estimation by K-mer distribution (K = 21). The x-axis represents the K-mer depth and the y-axis is the corresponding frequency. (**c**) Heatmap of Hi-C contact frequency across the *S. procerus* genome. (**d**) Circos plot of the *S. procerus* genome assembly.

In this study, a high-quality genome of *S. procerus* was *de novo* assembled from low-input genomic DNA of a single individual, using an ultra-low input approach and Hi-C (High-throughput chromosome conformation capture) technology. This genome provides insights into the evolutionary and genetic mechanisms behind *S. procerus*'s ecological adaptability and its interactions with host plants and natural enemies. It also lays a foundation for sustainable pest management strategies, including identifying adaptable strains and developing genetic tools for biological control.

Methods

Insect. *S. procerus* was originally collected in Hangzhou, Zhejiang Province, China, in 2024. The population was reared on water bamboos in a walk-in incubator at 26 °C under a photoperiod of 16:8 hours (light:dark), and relative humidity of 65%.

Genome survey. Before genome assembly, the genome size, heterozygosity, and repeat content were estimated using a Kmer-based method with Illumina sequencing data in Jellyfish v2.3.0[9], and the results were further

	Statistics
Number of scaffolds	334
Number of contigs	845
Genome size (bp)	752,845,220
Scaffold N50 (bp)	63,548,063
Contig N50 (bp)	1,882,825
Chromosome/total	97.52%
GC content (%)	35.28
Protein-coding gene	25,219

Table 1. Summary of the *S. procerus* genome assembly.

Pseudomolecule	Length (bp)
chr1	176,275,713
chr2	85,881,588
chr3	63,842,124
chr4	63,548,063
Chr5	53,745,909
chr6	51,573,254
chr7	46,143,111
chr8	41,918,599
chr9	40,905,903
chr10	32,382,405
chr11	31,225,671
chr12	23,788,616
chr13	13,638,171
chr14	9,423,431

Table 2. Pseudo-chromosome length of *S. procerus* by Hi-C.

analyzed with GenomeScope v2.015[10] (Fig. 1b), yielding a genome size of *S. procerus* 762.19 Mb and a heterozygosity rate of 1.11%.

Ultra-low input DNA sequencing. The HiFi library was constructed as our previous report[11] Briefly, a single adult male was washed with 75% ethanol and PBS, and then cleaned by removing the gut. Genomic DNA (gDNA) was extracted using the FineOut animal tissue DNA kit and its integrity was confirmed using the Femto Pulse system. The gDNA was fragmented using the Megaruptor 3 system and purified with ProNex beads. After PCR amplification, the SMRTbell library was constructed and size-selected with the BluePippin system, aiming for a 10 Kb insert size. The final 84.21 Gb of HiFi reads were used for de novo genome assembly.

To construct a high-quality genome assembly, PacBio HiFi reads were firstly converted from BAM to FASTA format using the bam2fasta tool in SMRT Link (v10.1)[12] with the parameters -c 1 -o prefix. Genome assembly was then performed using hifiasm v0.19.5[13] with default settings, generating an initial draft assembly. Redundancy in the assembly was minimized using purge_dups v1.2.5[14] with the parameters −2 -T. The final assembly resulted in a genome size of 765.58 Mb with a contig N50 of 1.16 Mb. The completeness of the assembled genome was assessed using BUSCO v5.4.4[15], which confirmed its high quality and completeness for downstream analyses.

Hi-C library construction and sequencing. For Hi-C sequencing and scaffolding, a pooled sample of 30 fifth-instar nymphs was crosslinked and quenched as described in our previous study[11]. Hi-C libraries were then constructed and sequenced on the Illumina NovaSeq X Plus platform. The draft genome was assembled to the chromosomal level using Hi-C data with Juicer v1.6.20[16] (https://github.com/aidenlab/juicer.) and Yahs v1.21[17] (https://github.com/c-zhou/yahs.). Visualization of contigs and scaffolds was performed using Juicebox v1.11.08[18]. The final *S. procerus* genome assembly measured 752.84 Mb with a contig N50 of 1.88 Mb (Table 1), with 97.52% of contigs anchored to 14 pseudo-chromosomes (Fig. 1c, d; Table 2).

Transcriptome sequencing. For assisting gene annotation, we prepared transcriptomes of *S. procerus* from eggs, nymphs (1st, 2nd, 3rd, 4th, and 5th), adult females and males. In addition, seven representative tissues including antennae, head, fat body, digestive gut, salivary gland, ovary, and testis were collected for RNA-sequencing (RNA-seq). RNA extraction and sequencing were performed as described in our previous study[11], and the obtained reads (Table 3) were used for subsequent genome annotation.

Gene prediction and annotation. RepeatMasker v4.1.2[19] with default parameters (http://repeatmasker.org/RepeatMasker) was applied to identify the repetitive sequences, encompassing interspersed repeats and

Filename	Reads number (Mb)	Bases (Gb)	Q30%
1st_nymph	40.0	6.0	95.19
2nd_nymph	40.0	5.9	94.58
3rd_nymph	38.7	5.8	94.20
4th_nymph	40.3	6.0	94.32
5th_nymph	41.1	6.2	95.02
Male	39.1	5.9	95.52
Female	39.6	5.9	94.62
Fat body	40.0	6.0	95.14
Antennae	46.5	7.0	95.95
Digestive gut	40.0	6.0	95.70
Head	40.0	6.0	95.31
Ovary	40.7	6.1	95.51
Salivary gland	43.5	6.5	95.16
Testis	40.0	6.0	95.86

Table 3. Summary of RNA-Seq data in this study.

Type	Number of elements	Length (bp)	Percentage (%)
DNAs	146,804,882	168,297,294	21.35
LTRs	249,340	54,140,966	7.2
LINEs	203,510	38,565,545	5.12
SINEs	1,022	104,848	0.01
Simple repeat	889	78,322	0.01
Others	201,179	56,185,458	7.46
Unknown	129,874	21,467,929	2.85
Total repeats	1,608,933	310,784,401	41.28

Table 4. Annotation of repeat elements in the *S. procerus* genome. Abbreviations: LTRs, long terminal repeats; LINEs, long interspersed nuclear elements; DNAs, DNA elements; SINEs, short interspersed nuclear elements.

transposable elements (TEs). Then RepeatModeler v2.0.3[20] with default parameters (http://www.repeatmasker.org/RepeatModeler) and Repbase library[21] (https://www.girinst.org/repbase) assisted to cluster the repeats by building a de novo repeat library. A total of 1,608,933 bp (41.28% of the genome) repetitive elements were identified, of which 33.22% was TEs, including long interspersed nuclear elements (LINEs) (5.12%), long terminal repeats (LTRs) (7.20%), short interspersed nuclear elements (SINEs) (0.01%), and DNA elements (DNAs) (21.35%) (Table 4).

Protein-coding gene annotation was performed using a combination of homology-based, transcriptome-based, and ab initio prediction methods. First, we used homologs from the selected Delphacidae species (*N. lugens*, *Sogatella furcifera*, and *Laodelphax striatellus*) as protein-based evidence for predicting gene sets using Miniprot v0.12[22]. RNA-seq reads were mapped using HISAT2 v2.2.1[23], and ab initio prediction was conducted using AUGUSTUS v3.5.0[24], trained with the transcriptome data. Finally, Protein-coding gene annotations from the three strategies was integrated using EvidenceModeler v2.1.0[25] (parameters:–segmentSize 800000–overlapSize 20000) to produce the final consensus gene set. This integrated annotation resulted in 24,349 protein-coding genes, and a BUSCO completeness score of 90.2% complete. Functional annotation of the predicted gene models was performed by searching against several databases, including Swiss-Prot, Nr (ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nr.gz), eggNOG[26], Pfam, GO, and KEGG. The result showed that 94.13% of the protein-coding genes had significant hits in the functional annotation database (Table 5).

Synteny analysis. For synteny analysis, genome data for *N. lugens*, *S. furcifera*, *L. striatellus* and *N. muiri* were obtained from NCBI. Synteny was analyzed using MCScanX v2[27] with default parameters, and the results were visualized with TBtools[28]. The analysis revealed high synteny between *S. procerus* and *N. muiri*, with 12,956 collinear genes identified, yielding a collinearity rate of 27.42%. This demonstrated strong genomic conservation across the Delphacidae family, highlighting the stability of their genome structure (Fig. 2a).

Orthologue and phylogenetic analyses. The protein sequences of 15 other Hemiptera species, including *N. muiri*, *N. lugens*, *S. furcifera*, *L. striatellus*, *Riptortus pedestris*, *Pyrrhocoris apterus*, *Oncopeltus fasciatus*, *Rhodnius prolixus*, *Cimex lectularius*, *Gerris buenoi*, *Myzus persica*, *Acyrthosiphon pisum*, *Aphis gossypii*, *Phenacoccus solenopsis*, and *Diaphorina citri*, were downloaded from InsectBase 2.0 (http://v2.insect-genome.com/) and figshare. Only the longest transcript of each gene was retained to represent the coding gene. OrthoFinder v2.5.5[29] with default parameters was used to determine the orthologs and paralogs, yielding 29,1499 orthologous gene

Database	Number	Percent (%)
Swiss-Prot	15402	46.64
GO	15108	44.33
KO	8718	24.27
NR	22738	67.22
PFAM	20991	73.73
eggNOG	11685	34.18
Annotated	20259	83 .20
Unannotated	4049	16.80

Table 5. Statistical analysis of the functional gene annotations of the *S. procerus* genome.

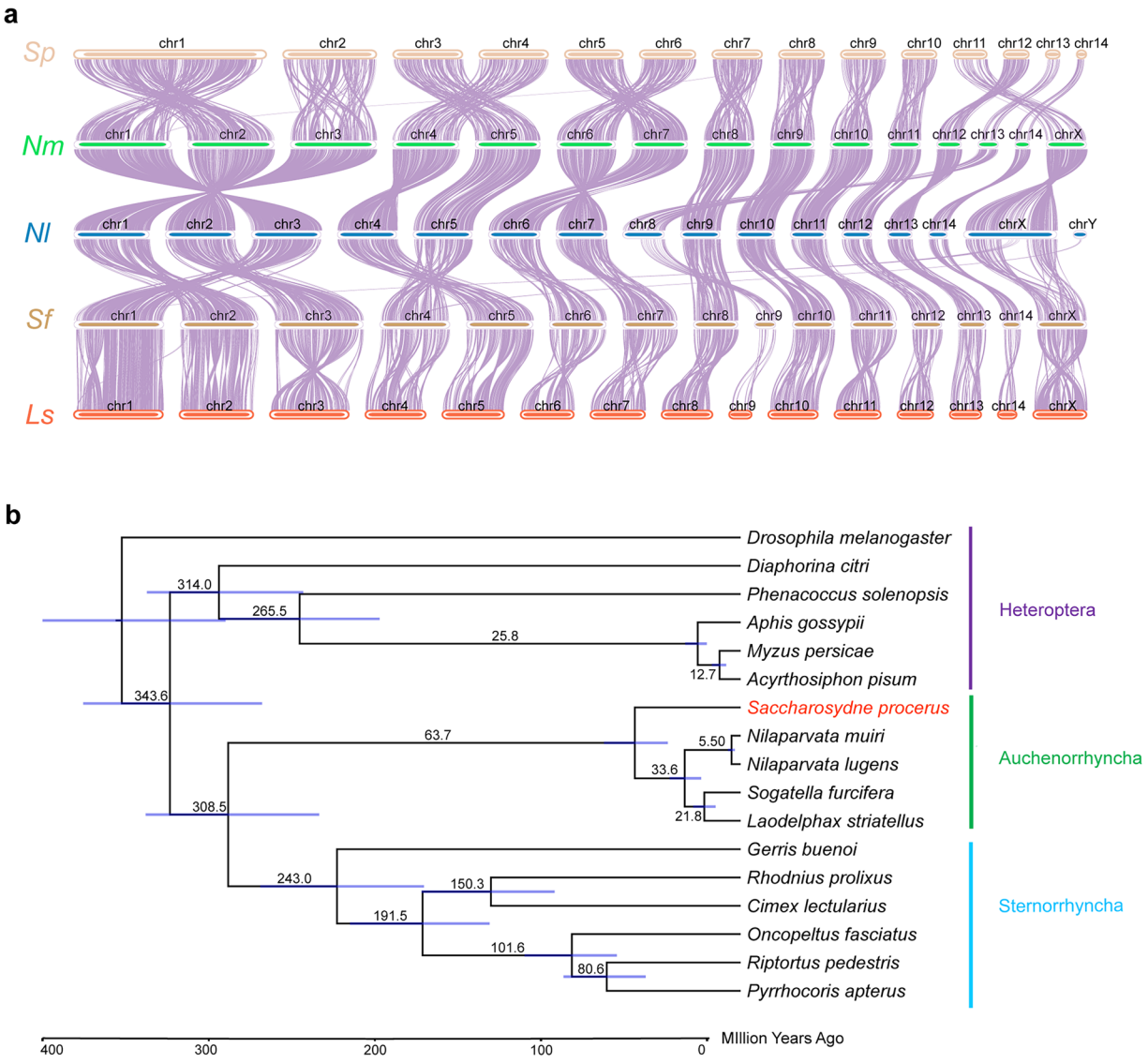

Fig. 2 Chromosomal synteny and phylogenetic analysis. (**a**) Chromosomal synteny among Delphacidae species: *Sp* (*S. procerus*), *Nm* (*N. muiri*), *Nl* (*N. lugens*), *Sf* (*S. furcifera*), and *Ls* (*L. striatellus*). Homologous chromosomal regions are connected by lines. (**b**) Maximum likelihood phylogenetic tree of *S. procerus* and 15 other Hemiptera species. The tree was constructed based on protein sequences, and divergence times (Mya) are indicated at each node. All nodes are supported by 1,000 bootstrap replicates (values shown as percentages). The scale bar at the bottom represents divergence time.

families from a total of 341,893 genes. All identified 242 single-copy orthologs were aligned using MAFFT v7.520[30] and trimmed utilizing trimAl v1.4.1[31]. The maximum likelihood tree was constructed using IQ-TREE v2.2.3[32] with 1,000 ultrafast bootstrap replicates, and D. melanogaster was used as an outgroup. The divergence

time between different species was estimated using MCMCtree (PAML v4.10.7 package)[33] based on the fossil records acquired from the TimeTree database[34] (http://www.timetree.org/) using the approximate likelihood calculation method (*A. gossypii* vs. *A. pisum* 23.0–64.5 million years ago (MYA), *D. citri*vs. *A. pisum* 211.9–351.0 MYA, and *R. prolixus* vs. C. *lectularius* 108.0–212.9 MYA). The results of the phylogenetic analysis revealed that *S. procerus* forms a distinct clade within the suborder *Auchenorrhyncha*, with an estimated divergence time of approximately 63.7 million years ago (MYA) (Fig. 2b).

Data Records

The genome assembly, which utilized PacBio and Hi-C sequencing data, along with the transcriptome data used for genome annotation, has been stored in the NCBI Sequence Read Archive under accession numbers SRR32281657 and SRR32281643, with BioProject accession numbers PRJNA1221133[35]. The genome assembly has been deposited at DDBJ/ENA/GenBank under the accession JBLNNE000000000[36]. The genome annotation files are available in Figshare under a DOI number of https://doi.org/10.6084/m9.figshare.28375412[37].

Technical Validation

The final genome assembly of *S. procerus* was 752.84 Mb with a contig N50 of 1.88 Mb, and 97.52% contigs were anchored to 14 pseudo-chromosomes. The BUSCO results showed that 94.8% of the complete Insecta genes were captured by the genome. Additionally, the gene prediction result was assessed by BUSCO, which indicated that 90.2% genes (single-copied gene: 89%, and duplicated gene: 1.2%) were complete. Furthermore, 95.57% of the Illumina clean reads were accurately mapped back to the assembly. A high mapping success rate of 95.19% was also achieved for the RNA sequencing reads. Overall, these assessments reflect the high quality of the genomic assembly.

Code availability

Software for data analyses was mentioned in Methods. The core code and parameters are available at https://github.com/zhuozhuoliu/Sp.genome.

Received: 2 March 2025; Accepted: 16 June 2025;
Published online: 01 July 2025

References

1. Chen, Y. Y. *et al.* Abundant genetic diversity of the wild rice *Zizania latifolia* in central China revealed by microsatellites. *Ann. Appl. Biol.* **161**, 192–201 (2012).
2. Chen, Y. Y. *et al.* Landscape-scale genetic structure of wlid rice *Zizania latifolia*: the roles of rivers, mountains and fragmentation. *Front. Ecol. Evol.* **5**, 17 (2017).
3. Chen, J. *Identification and ecological control of insect pests of wateroat*. (China Agriculture Press. pp. 1–168 2016).
4. Shan, L. *et al.* Genetic diversity analysis of *Saccharosydne procerus* in Hunan region, China. *Mol. Biol. Rep.* **51**, 878 (2024).
5. Yin, H. C. *et al.* Effects of parental body color and environmental factors on the body color, longevity and reproduction of *Saccharosydne procerus* (Hemiptera: Delphacidae) adults. *Acta Entomologica Sinica.* **64**, 1197–1204 (2019).
6. Zhong, H. *et al.* The luring effect of different sticky traps on the *Saccharosydne procerus* (Matsumura). *J. Zhejiang Agric. Sci.* **57**, 1642–1643 (2016).
7. Zheng, X. S. *et al.* Migration patterns of natural enemies of rice planthoppers between rice paddies and paddy fields. *Acta Agriculturae Zhejiangensis* **6**, 339–343 (1999).
8. Qing, Y. X. *et al.* A study on the relationship between egg parasitoid, *anagrus nilaparvatae* and green slender planthopper, *Saccharosydne procerus*, a species of insect pest of wild rice, *Zizania Caduciflora*. *Acta Entomologica Sinica* **42**, 387–393 (1999).
9. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
10. Vurture, G. W. *et al.* GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
11. Liu, Z. Q. *et al.* Chromosome-level genome assembly from a single planthopper *Nilaparvata muiri* (Hemiptera: Delphacidae). *Scientific Data* **11**(1), 937 (2014).
12. Rhoads, A. & Au, K. F. PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics* **13**, 278–289, https://doi.org/10.1016/j.gpb.2015.08.002 (2015).
13. Cheng, H. *et al.* H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods* **18**, 170–175 (2021).
14. Guan, D. *et al.* Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinforma. Oxf. Engl.* **36**, 2896–2898 (2020).
15. Seppey, M., Manni, M. & Zdobnov, E. M. BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods Mol Biol.* **1962**, 227–245 (2019).
16. Durand, N. C. *et al.* Juicer provides a One-Click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
17. Zhou, C., McCarthy, S. A. & Durbin, R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics.* **39**, btac808 (2022).
18. Robinson, J. T. *et al.* Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data. *Cell Syst.* **6**, 256–258.e1 (2018).
19. Smit, A., Hubley, R. & Green, P. RepeatMasker Open-4.0. http://www.repeatmasker.org (2015).
20. Smit, A., & Hubley, R. RepeatModeler. http://www.repeatmasker.org/RepeatModeler/ (2008–2015).
21. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 1–6 (2015).
22. Li, H. Protein-to-genome alignment with miniprot. *Bioinformatics* **39**, btad014 (2023).
23. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. methods* **12**, 357–360 (2015).
24. Stanke, M. *et al.* AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Res.* **34**, 435–439 (2006).
25. Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology* **9**, R7 (2008).
26. Huerta-Cepas, J. *et al.* EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, 309–314 (2019).

27. Wang, Y. *et al*. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, 1–14 (2012).
28. Chen, C. *et al*. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant* **13**, 1194–1202 (2020).
29. Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 1–14 (2019).
30. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
31. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
32. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
33. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
34. Kumar, S., Stecher, G., Suleski, M. & Blair Hedges, S. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
35. Liu, Z. Q. *et al*. Whole genome sequencing of *Saccharosydne procerus*. *NCBI Sequence Read Archive*. https://identifiers.org/ncbi/insdc.sra:SRP562622 (2025).
36. Liu, Z. Q. *et al*. Chromosome-level genome assembly of *Saccharosydne procerus*. *GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_048001795.1 (2025).
37. Liu, zhuoqi Genome annotation file of the *Saccharosydne procerus*. *figshare. Dataset.* https://doi.org/10.6084/m9.figshare.28375412.v1 (2025).

Acknowledgements

This research was supported by the National Key Research and Development Program of China (2023YFF1000500), and National Natural Science Foundation of China (32272519).

Author contributions

H.J.X. and J.F.Z. convinced and designed the study. J.F.Z. reared the S. procerus, J.F.Z., Z.Q.L., L.X.Y., N.T.X. and J.L. collected the samples. X.Z. and Z.Q.L. performed bioinformatics analysis. Z.Q.L. wrote the draft manuscript. H.J.X. and J.F.Z. modified 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 H.-J.X. or J.-F.Z.