scientific data

OPEN

DATA DESCRIPTOR

A telomere-to-telomere genome assembly for *Cyperus difformis*

Jingbo Li[1,5], Jian Zhao[2,5], Wei Zheng[3,5], Yonglin Ma[4], Meisha Qiu[4], Weisheng Li[4] & Yanhui Wang[3,4] ✉

Cyperus difformis **is a globally problematic weed in rice fields, posing a significant threat to rice yield. While chemical herbicides are commonly used for its control, the species often escapes management due to its rapid evolution of resistance to widely used herbicides. To better understand the mechanisms underlying herbicide resistance, insights into the genetics and genomics of *C. difformis* are essential. In this study, we present a telomere-to-telomere genome assembly of *C. difformis*, generated by combining PacBio HiFi, Oxford Nanopore Technologies (ONT), MGI short reads and high-throughput chromatin conformation capture (Hi-C) technologies. The assembled genome spans 226 Mb with a scaffold N50 of 13.08 Mb. Utilizing Hi-C interaction data, 97.24% of the contigs were anchored to 18 chromosomes, with 35 telomeres successfully defined. Further analysis identified 75.73 Mb of repetitive sequences and 21,069 protein-coding genes, of which 91.8% (19,347 genes) were functionally annotated. This high-quality genome provides a valuable resource for studies in population genetics, phylogeny, comparative genomics, adaptive evolution, and functional genomics of *C. difformis*.**

Background & Summary

Cyperus difformis, native to southern Europe, Africa, Asia, and Australia, has become an invasive species in many parts of the world[1]. It is recognized as a major weed in rice fields across 46 countries[2], thriving in fertile soils while also adapting to less favorable sandy or clay environments[1,3]. As a highly self-pollinating annual weed, *C. difformis* produces abundant seeds and can complete its life cycle in as little as one month[4]. This rapid development gives it a competitive advantage over slower-growing crops like rice, which typically require around three months to mature[1]. The weed competes aggressively with rice from early growth stages and continues through critical yield-determining periods, significantly reducing panicle numbers[5]. In China, the threat posed by *C. difformis* has intensified following a shift from traditional puddled-transplanted rice cultivation to direct-seeding practices. In these systems, *C. difformis* has emerged as one of the most problematic weeds[2,3].

Currently, utilization of chemical herbicides is the dominant method to control this weed in rice fields. Commonly used herbicides include pretilachlor, pyrazosulfuron-ethyl, bensulfuron methyl, MCPA, bentazone, pyribenzoxim, penoxsulam, butachlor, bispyribac-sodium and halosulfuron methyl[4]. However, the long-term pressure exerted by herbicides on this weed have extensively incurred herbicide resistance across multiple modes of action. Resistance has been documented against ALS inhibitors (e.g., bensulfuronmethyl, azimsulfuron, halosulfuron-methyl, ethoxysulfuron, imazamox, pyrazosulfuron-ethyl, imazosulfuron, cinosulfuron, cyclosulfamuron, penoxsulam, and imazapic) and ACCase inhibitors (e.g., bispyribac-sodium)[6–8]. Cases of herbicide-resistant *C. difformis* have been reported in Australia, Italy, Brazil, Spain, South Korea, Greece, Turkey, the United States, India and China[8]. The rapid evolution of resistance in this species involves both target- and non-target-resistance mechanisms[4,6,7,9–13], complicating the identification of underlying genetic and physiological factors.

A telomere-to-telomere genome assembly for *C. difformis* would provide critical insights into the evolution of herbicide resistance, enabling the identification of resistance-associated genes and regulatory networks. In this

[1]Engineering Research Center for Salt- and Alkali-Tolerant Food Crop Seeds of Guangxi Zhuang Autonomous Region, Guangxi Vocational University of Agriculture, Nanning, 530000, China. [2]School of Tropical Agriculture and Forestry, Hainan University, Haikou, 570228, China. [3]Guangxi Crop Genetic Improvement and Biotechnology Laboratory, Guangxi Academy of Agricultural Sciences, Nanning, 530007, China. [4]Key Laboratory of Green Prevention and Control on Fruits and Vegetables in South China Ministry of Agriculture and Rural Affairs, Guangxi Key Laboratory of Biology for Crop Diseases and Insect Pests, Plant Protection Research Institute, Guangxi Academy of Agricultural Sciences, Nanning, 530007, China. [5]These authors contributed equally: Jingbo Li, Jian Zhao, Wei Zheng. ✉e-mail: wangyh1984@163.com

Types	Sample	Platform	Bases (Gb)	Reads Count	Mean length (bp)	N50 (bp)
ONT	Leaf	ONT promethION	27.3	777,767	35,217.79	100,000
HiFi	Leaf	PacBio Revio	21.2	1,140,966	18,572.40	17,863
Hi-C	Leaf	MGI DNBSEQ-T7	23.6	78,675,058	150	150
Short-reads	Leaf	MGI DNBSEQ-T7	27.8	92,787,005	150	150
RNA-seq	Leaf	MGI DNBSEQ-T7	13.4	44,623,133	150	150

Table 1. Sequencing data for *C. difformis*.

study, the telomere-to-telomere genome of *C. difformis*, totaling 219.76 Mb, was successfully assembled through an integrative analysis of PacBio HiFi and ONT, NGS and Hi-C. A total of 35 telomeres were identified across 18 chromosomes. The assembled genome contained 75.73 Mb of repetitive sequences, accounting for 33.51% of the total genome. 21,069 protein-coding genes were identified with mean exon and CDS lengths of 220.07 bp and 1,205.97 bp, respectively. Functional annotation revealed that 91.8% of these predicted genes were supported by existing public databases. Additionally, a comprehensive set of non-coding RNAs (ncRNAs) was identified, including 328 miRNAs, 2,326 tRNAs, 2,553 rRNAs, and 269 snRNAs. Such genomic resources could inform strategies to mitigate resistance development and enhance sustainable weed management in agroecosystems.

Methods

Sample collection. *A C. Difformis* population was collected from a rice field in Guigang city, Guangxi Zhuang Autonomous Region (23°18′51″N, 109°55′55″E)[12]. a individual of this population was used for genome sequencing.

DNA extraction and genome sequencing. Genomic DNA was extracted from leaves using the CTAB method[14]. Quality and quantity of DNA were assessed using agarose gel electrophoresis and an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA).

For long-read sequencing, one library was constructed using SQK-LSK109 Kit (Oxford Nanopore Technologies, Oxford, UK), and sequenced on the ONT promethION System. Another library was constructed using the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, Menlo Park, CA, USA), and sequenced on the PacBio Revio System. This generated a total of 27.3 Gb of ONT reads with an N50 of 100,000 bp and 21.2 Gb of HiFi reads with an N50 of 17,863 bp, using the Circular Consensus Sequencing (CCS) software v6.0.0.25 with the optimized parameter set to -min-passes 3 (Table 1).

For short-read sequencing, a library with a 350 bp insert size was prepared using NEBNext Ultra DNA Library Prep Kit (New England Biolabs, Ipswich, MA, USA). Hi-C sequencing was performed using a library constructed with the GrandOmics Hi-C Kit (GrandOmics, Wuhan, China), in which DpnII enzyme was used to digest genomic DNA. Both the short-read and Hi-C libraries were sequenced on the MGI DNBSEQ-T7 platform. In total, 27.8 Gb of short reads and 23.6 Gb of Hi-C reads were obtained (Table 1).

RNA extraction and RNA-seq. Total RNA was obtained using the MiniBEST Plant RNA Extraction Kit (TaKaRa Biotechnology, Beijing, China). The quality and integrity of the RNA were assessed using a NanoDroop 2000c (Thermo Fisher Scientific, Waltham, MA, USA) and an Agilent 2100 Bioanalyzer (Agilent Technologies), respectively. RNA samples with OD260/280 $\geq$ 1.8 and RNA integrity number $\geq$ 7.0 were selected for transcriptome sequencing. The cDNA libraries were then prepared and sequenced on the platform of MGI DNBSEQ-T7. Average approximately 13.4 Gb of transcriptome sequencing data were generated for assistance to gene structure prediction (Table 1).

Genome size estimation. The genomic characteristics of *C. difformis* were assessed using a K-mer analysis. K-mer (K = 17) counts were generated with Jellyfish (v2.2.6)[15], and the results were analyzed using GenomeScope v2.0[16]. The genome was estimated to be approximately 204.58 Mb in size, with a repeat content of 33.1% and a heterozygosity rate of 0.793% (Fig. 1).

Genome assembly. PacBio sequencing data were filtered to remove low-quality polymerase reads, and the remaining subreads were processed using SMRT Link v8.0 with parameters --min-passes = 3 and --min-rq = 0.99. Through this process, 27.3 Gb of ONT reads and 21.2 Gb of HiFi reads were produced with mean read lengths of 35.21 kb and 18.57 kb, respectively (Table 1). The resulting long reads were assembled using HiFiasm 0.19.9-r616[17] with default parameters, producing a scaffold-level genome assembly of 226 Mb in total length, comprising 96 scaffold with a scaffold N50 of 13.08 Mb (Table 2).

Chromosome construction. Hi-C clean reads were mapped to the assembled contigs using Bowtie2 v2.2.4[18]. Valid ligation products were then identified using HiCUP v0.8.0[19], and only valid paired-end reads were retained for downstream analysis. These valid contacts were used to cluster, order, and orient contigs into chromosomes using LACHESIS[20] with optimized parameters: CLUSTER_MIN_RE_SITES = 50, CLUSTER_NONINFORMATIVE_RATIO = 0, and CLUSTER_MAX_LINK_DENSITY = 3. Assembly candidates were visualized and manually curated using Juicebox v1.11.08[21]. As a result, we obtained a final telomere-to-telomere genome assembly totaling 219.76 Mb, with 97.24% of the assembled contigs anchored to 18 chromosomes. A total of 35 telomeres were identified across these chromosomes based on the conserved telomeric repeat sequence "CCCTAAA" (Figs. 2 and 3, Table 3).

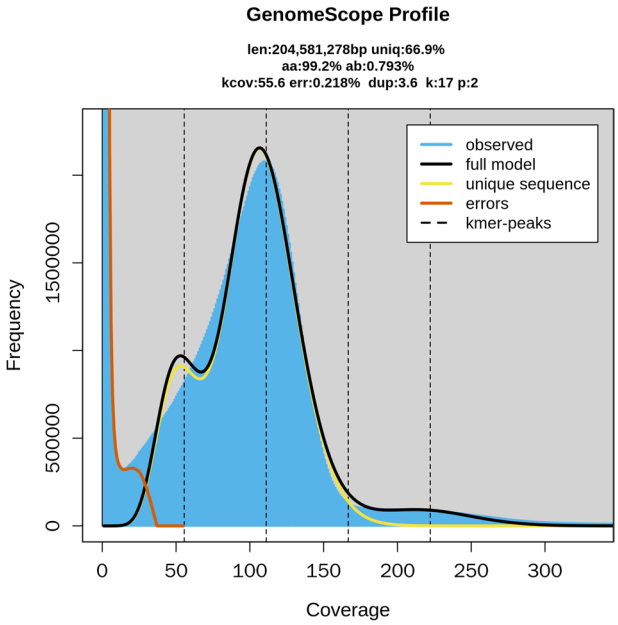

Fig. 1 The K-mer (k = 17) distribution and genome size estimation of *C. differmis* genome.

Statistic type	Contig	Scaffold
N50	13,081,375	13,081,375
N90	9,411,674	9,411,674
L50	8	8
L90	17	17
Shortest Length	25,337	25,337
Longest Length	17,211,812	17,211,812
Total length	225,999,259	225,999,259
Number of Fragments	96	96
GC Content (%)	33.89%	33.89%

Table 2. Genome assembly statistics for *C. differmis*.

The completeness of the assembled genome was assessed using BUSCO v5.2.2[22], based on the *embryophyta_odb10* lineage dataset. The analysis identified 92.6% as single-copy BUSCOs, 1.6% as duplicated BUSCOs, 1.9% as fragmented BUSCOs, and 3.9% as missing BUSCOs (Table 4). In addition, genome quality was evaluated using Merqury[23], which yielded a QV of 42.92 based on short-read data with a K-mer size of 21 (Table 4). These results collectively indicate that the assembled genome possesses high completeness, continuity, and accuracy, qualifying it as a high-quality reference genome.

Repeat sequences annotation. Repeat and LTR sequences in the *C. differmis* genome were predicted de novo using RepeatModeler v1.0.3[24], RepeatScout v1.0.5[25] and LTR_Finder v1.07[26]. The predicted LTR sequences were deduplicated using LTR_retriever v2.9.0 (parameter: -threads 16) and then merged with detected repeat elements and the RepBase library (v20181026) to create the repeat library. Repetitive elements were identified using RepeatMasker v4.0.9[27] (parameters: -nolow -no_is -norna -parallel 2) and RepeatProteinMask v4.0.9 (parameters: -noLowSimple -pvalue 0.0001) against the repeat library. After merging and deduplicating the results, we identified a total of 75.73 Mb of repeat sequences, which accounted for 33.51% of the assembled genome (Table 5).

Structural and functional annotation of protein coding genes. To annotate protein-coding genes in the *C. differmis* genome, an integrated approach combining transcriptome-based, homology-based, and ab initio predictions was employed.

For transcriptome-based prediction, raw RNA-seq data were first filtered using fastp v0.21.0[28], then aligned to the genome with TopHat2 v2.2.4[29], and assembled into transcripts using Trinity v2.1.1[30]. The resulting transcripts were mapped to the genomes using the Program to Assemble Spliced Alignment[31] (v2.5.3, PASA, https://github.com/PASApipeline/PASApipeline) and obtained coding regions.

For homology-based prediction, protein sequences from six reference species—*Arabidopsis thaliana*, *Carex littledalei*, *Oryza sativa*, *Rhynchospora breviuscula*, *Rhynchospora pubera*, and *Rhynchospora tenuis*—were aligned to the genome using TBLASTN[32] (BLASTall v2.2.26) with an e-value cut-off of 1×10^{-5}. The Basic Local

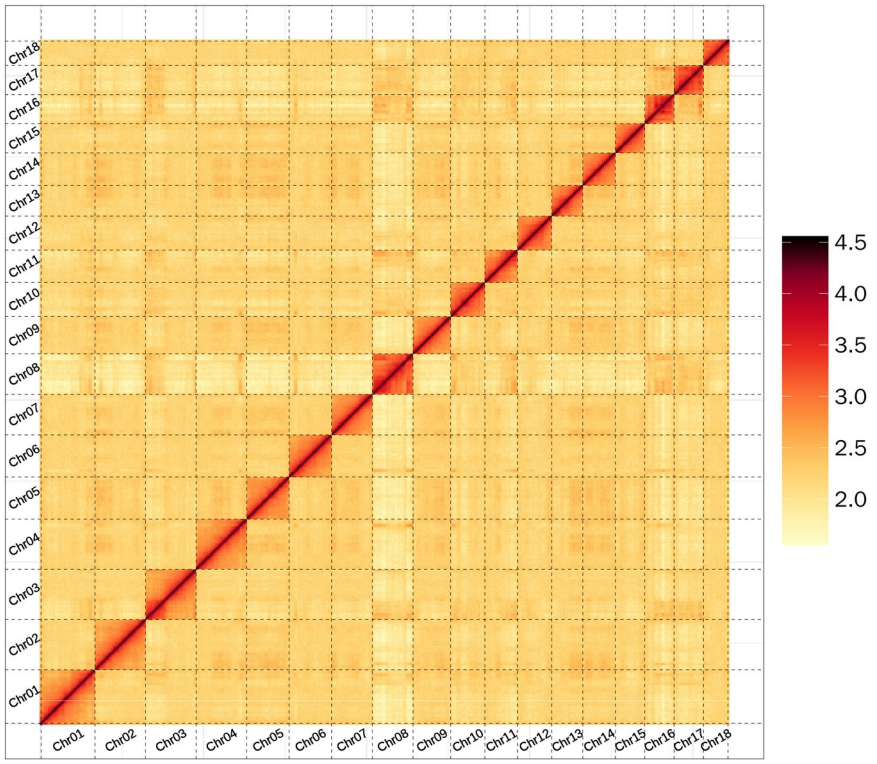

Fig. 2 The genome-wide chromosomal interactions heatmap based on Hi-C Data.

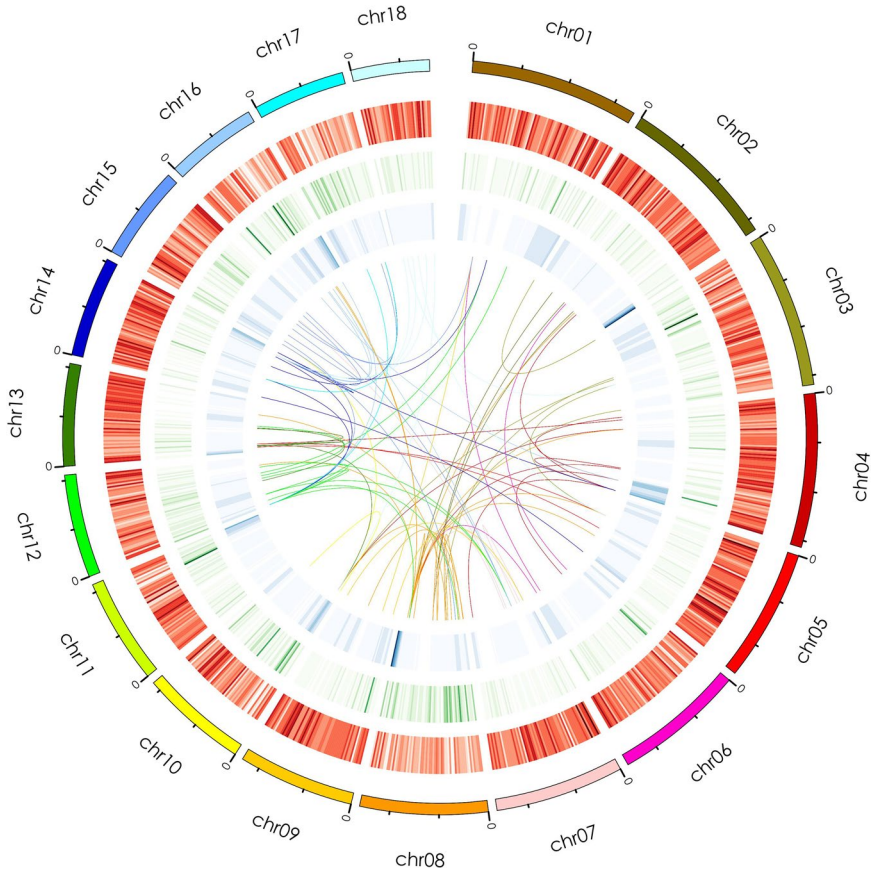

Fig. 3 Circos ideogram of the *C. diffformis* genome. Tracks from outer to inner include 18 chromosomes, gene density, repeat element density, GC content, and > 10 Kbp genome fragment duplication.

Chromosome	Length (bp)	Percentage (%)	Gene number	Start of left telomere	End of left telomere	Length of Left telomere	Start of right telomere	End of right telomere	Length of right telomere
Chr01	17,211,812	7.62	1,605	1	3841	3841	17209426	17211812	2387
Chr02	16,207,434	7.17	1,656	1	4388	4388	16204746	16207434	2689
Chr03	15,951,341	7.06	1,459	1	4790	4790	15946117	15951341	5225
Chr04	15,713,177	6.95	1,636	1	6190	6190	15708916	15713177	4262
Chr05	13,548,089	5.99	1,358	1	439	439	13545719	13548089	2371
Chr06	13,376,879	5.92	1,352	1	4775	4775	13372291	13376879	4589
Chr07	13,180,414	5.83	1,268	1	3935	3935	13175997	13180414	4418
Chr08	13,081,375	5.79	990	1	4241	4241	NA	NA	NA
Chr09	11,892,291	5.26	1,227	1	4511	4511	11887447	11892291	4845
Chr10	10,981,906	4.86	947	1	3700	3700	10975431	10981906	6476
Chr11	10,806,829	4.78	959	1	6430	6430	10802739	10806829	4091
Chr12	10,564,411	4.67	961	1	5499	5499	10558407	10564411	6005
Chr13	10,426,930	4.61	1,123	1	4871	4871	10423283	10426930	3648
Chr14	10,309,465	4.56	1,144	1	4708	4708	10301741	10309465	7725
Chr15	9,642,368	4.27	911	1	4739	4739	9634966	9642368	7403
Chr16	9,428,065	4.17	762	1	4759	4759	9422393	9428065	5673
Chr17	9,411,674	4.16	738	1	4653	4653	9407103	9411674	4572
Chr18	8,024,608	3.55	734	1	5830	5830	8024252	8024608	357
Total	219,759,068	97.24	20,830	–	–	–	–	–	–
Unplaced	6,240,191	2.76	239	–	–	–	–	–	–

Table 3. Summary of chromosome information of the *C. difformis*.

Type	Genome		Annotation	
	Number	Percentage (%)	Proteins	Percentage (%)
Complete BUSCOs (C)	1521	94.2	1492	92.4
Complete and single-copy BUSCOs (S)	1495	92.6	1464	90.7
Complete and duplicated BUSCOs (D)	26	1.6	28	1.7
Fragmented BUSCOs (F)	30	1.9	47	2.9
Missing BUSCOs (M)	63	3.9	75	4.7
QV	42.92	–	–	–

Table 4. The BUSCO integrity, consistency and consensus quality value (QV) evaluation of the genome and annotation.

Type	TE protiens		Repbase + Denovo		Combined TEs	
	Length (bp)	% in genome	Length (bp)	% in genome	Length (bp)	% in genome
DNA	416,001	0.18	10,096,762	4.47	10,156,628	4.49
LINE	30,406	0.01	1,394,751	0.62	1,408,139	0.62
SINE	0	0.00	1,176,364	0.52	1,176,364	0.52
LTR	3,731,295	1.65	41,380,104	18.31	41,496,840	18.36
Other	0	0.00	155	0.00	155	0.00
Satellite	0	0.00	163,836	0.07	163,836	0.07
Simple_repeat	0	0.00	592,400	0.26	592,400	0.26
Unknown	0	0.00	28,362,672	12.55	28,362,672	12.55
Total	4,177,702	1.85	75,596,760	33.45	75,731,051	33.51

Table 5. Summary of the repetitive elements in the *C. difformis* genome.

Alignment Search Tool (BLAST) hits were joined together by Solar v0.9.6[33]. GeneWise v2.4.1[34] was used to predict the exact gene structure of the corresponding genomic regions for each BLAST hit.

For ab initio prediction, multiple tools were used on the repeat-masked genome: AUGUSTUS v3.2.3 (parameters: --uniqueGeneId = true --noInFrameStop = true --gff3 = on --strand = both)[35], Genscan[36], GlimmerHMM v3.0.4 (parameters: -f -g)[37], SNAP (v2013-11-29)[38], and GeneID v1.4[39].

All prediction results were integrated using EvidenceModeler v1.1.1 (EVM)[40]. The weights for each type of evidence were set as follows: PASA-set > Homolog-set > Cufflinks-set > Augustus > Geneid = SNAP = Glimm erHMM = GeneScan. The gene models were further updated by PASA2 to generate untranslated regions and

Method	Gene set	Gene Number	Average gene length (bp)	Average CDS length (bp)	Average exons per gene	Average exon length (bp)	Average intron length (bp)
Ab initio	Augustus	23,579	2611.82	1118.29	4.92	227.27	380.96
	GlimmerHMM	36,597	4349.78	777.07	3.46	224.73	1453.67
	SNAP	46,810	2322.13	728.42	4.96	146.91	402.61
	Geneid	28,250	4792.34	941.73	4.91	191.82	984.94
	Genscan	21,413	6543.02	1230.33	6.06	203.06	1050.16
Homology-based	Arabidopsis thaliana	3,190	1367.7	874.89	2.34	374.41	368.69
	Carex littledalei	11,342	2716.69	1168.79	5.27	221.73	362.4
	Oryza sativa	4,082	1552.03	886.07	2.83	312.59	363
	Rhynchospora breviuscula	12,059	2527.57	1126.62	4.89	230.29	359.94
	Rhynchospora pubera	15,125	2136.1	1014.1	4.3	235.87	340.07
	Rhynchospora tenuis	15,125	2136.1	1014.1	4.3	235.87	340.07
RNAseq	Cufflinks	40,069	4433.77	2121.64	6.53	324.7	417.79
	PASA	45,373	2630.86	1079.45	5.42	199.25	351.2
EVM		24,920	2747.67	1114.16	5.06	220.34	402.69
PASA-update		24,717	2730.53	1130.42	5.08	222.51	392.15
Final set		21,069	2953.15	1205.97	5.48	220.07	389.99

Table 6. Statistic of the predicated protein coding genes by different methods.

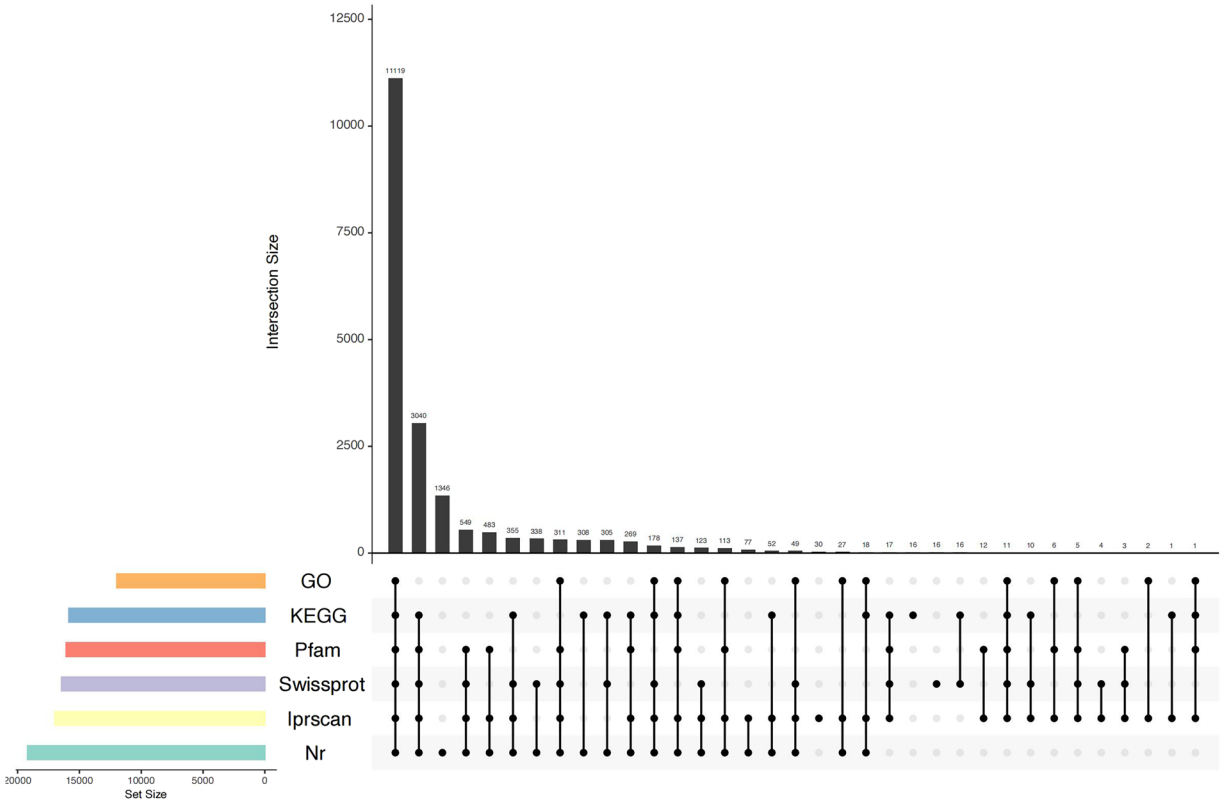

Fig. 4 The upset bar plot visualizing the functional annotation results based on different databases.

alternative splicing variation information. This process yielded a high-confidence set of 21,069 protein-coding genes. On average, each gene contained 5.48 exons, with mean exon and CDS lengths of 220.07 bp and 1,205.97 bp, respectively (Table 6).

Functional annotation of the predicted protein-coding genes was performed by aligning them to multiple public databases, including NR, Swiss-Prot, KEGG, InterPro, Pfam, and GO. Sequence alignments were conducted using Diamond v0.9.14 (parameter: --evalue 1e-5)[41]. Additionally, protein motifs, domains, and conserved sequences were identified using HMMER (Hmmscan v3.3.2, parameter: -E 0.01) and InterProScan v5.63-95 (parameters: -goterms -pa -dp -verbose)[42]. As a result, 91.8% (19,347) of the predicted genes were successfully annotated with functional information across public databases (Fig. 4, Table 7).

Item	Annotated Number of putative genes	Annotated Percentage
NR	19,197	91.1
Swissprot	16,449	78.1
KEGG	15,853	75.2
Interpro	17,002	80.7
Pfam	16,076	76.3
GO	11,977	56.8
Annotated	19,347	91.8
Total genes	21,069	100

Table 7. Summary of the functionally annotated genes in the genome of *C. difformis*.

Type		Copy(w)	Average length (bp)	Total length (bp)	% of genome
miRNA		328	119.9268	39,336	0.0174
tRNA		2,326	75.7481	176,190	0.0780
rRNA	rRNA	2,553	363.9373	929,132	0.4111
	18S	372	1739.9731	647,270	0.2864
	28S	1,426	144.0217	205,375	0.0909
	5.8S	356	117.7472	41,918	0.0185
	5S	399	86.6391	34,569	0.0153
snRNA		269	119.0818	32,033	0.0142

Table 8. Statistic of the identified non-coding RNAs in the genome.

In addition, ncRNAs were annotated using a combination of specialized tools. rRNAs were identified using RNAmmer v1.2 (parameters: -S euk -m tsu, lsu, ssu)[43], while tRNAs were predicted with tRNAscan-SE v1.3.1 (parameters: -E -j tRNA.gf -o tRNA.result -f tRNA.struct -thread 16)[44]. Other small ncRNAs, including microRNAs (miRNAs) and small nuclear RNAs (snRNAs), were annotated using INFERNAL v1.1.4 (parameters: -cut_ga -rfam -nohmmonly -fmt 2)[45]. In total, 328 miRNAs, 2,326 tRNAs, 2,553 rRNAs, and 269 snRNAs were identified in the *C. difformis* genome (Table 8).

Data Records

All data generated in this study are publicly available. The raw sequence data have been deposited in the Genome Sequence Archive (Genomics, Proteomics & Bioinformatics 2025) in National Genomics Data Center (Nucleic Acids Res 2025) under CRA031324[46], China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences. The telomere-to-telomere level assembly and annotation file of *Cyperus difformis* are available at Figshare[47]. The assembly file is also deposited in NCBI under JBTANI000000000[48].

Technical Validation

We performed two key assessments of the assembled genome: (i) genome completeness using BUSCO v5.2.2[22] with the embryophyta_odb10 lineage dataset; (ii) quality scoring using Merqury v1.3[23], based on NGS reads with a "k = 21" parameter. The genome completeness and consensus quality (QA) score of the CDL genome were 94.2% and 42.92, respectively, confirming the assembly as reference quality. Furthermore, the annotated proteins, evaluated with BUSCO using the embryophyta_odb10 dataset, achieved a score of 92.4%, highlighting the high quality of the annotation.

Data availability

The raw sequence data are available in the Genome Sequence Archive at https://ngdc.cncb.ac.cn/gsa/browse/CRA031324. The assembly and annotation file of *Cyperus difformis* are available from Figshare at https://doi.org/10.6084/m9.figshare.29072027. the assembly file is also available in NCBI at https://www.ncbi.nlm.nih.gov/nuccore/JBTANI000000000.

Code availability

All bioinformatic analyses were conducted in accordance with the respective software manuals and standard protocols. Software versions and critical parameters are reported in the Methods section, with unspecified settings retained as defaults. No custom code was developed during this study.

Received: 16 May 2025; Accepted: 8 January 2026;
Published online: 15 January 2026

References

1. Holm, L. G., Plucknett, D. L., Pancho, J. V. & Herberger J. P. The World's worst weeds: Distribution and biology. Malabar, FL, USA: UH Press 609 (1991).
2. Rao, A. N., Johnson, D. E., Sivaprasad, B., Ladha, J. K. & Mortimer, A. M. Weed Management in Direct-Seeded Rice. *Adv. Agron.* **93**, 153–255 (2007).
3. Chauhan, B. S. & Johnson, D. E. Ecological studies on *Cyperus difformis*, *Cyperus iria* and *Fimbristylis miliacea*: three troublesome annual sedge weeds of rice. *Ann. Appl. Biol.* **155**, 103–112 (2009).
4. Guo, X. *et al*. Effect of mutations on acetohydroxyacid synthase (AHAS) function in *Cyperus difformis* L. *J. Integr. Agric.* **23**, 177–186 (2024).
5. Swain, D. J., Nott, M. J. & Trounce, R. B. Competition between *Cyperus difformis* and rice: the effect of time of weed removal. *Weed Res.* **15**, 149–152 (1975).
6. Li, Z. *et al*. Variation in mutations providing resistance to acetohydroxyacid synthase inhibitors in *Cyperus difformis* in China. *Pestic. Biochem. Phys.* **166**, 104571 (2020).
7. Choudhary, V. K. *et al*. Resistance in smallflower umbrella sedge (*Cyperus difformis*) to an acetolactate synthase-inhibiting herbicide in rice: First case in India. *Weed Technol.* **35**, 710–717 (2021).
8. Heap I. The international herbicide-resistant weed database. http://www.weedscience.org (2025).
9. Ntoanidou, S., Kaloumenos, N., Diamantidis, G., Madesis, P. & Eleftherohorinos, I. Molecular basis of *Cyperus difformis* cross-resistance to ALS-inhibiting herbicides. *Pestic. Biochem. Phys.* **127**, 38–45 (2016).
10. Tehranchian, P. *et al*. ALS-resistant smallflower umbrella sedge (*Cyperus difformis*) in Arkansas rice: physiological and molecular basis of resistance. *Weed Sci.* **63**, 561–568 (2015).
11. Huang, M. *et al*. Comparative analysis of resistance to ALS-inhibiting herbicides in smallflower umbrella sedge (*Cyperus difformis*) populations from directseeded and puddled-transplanted rice systems. *Weed Sci.* **70**, 174–182 (2022).
12. Chen, X. *et al*. Multiple herbicide resistance in a *Cyperus difformis* population in rice field from China. *Pestic. Biochem. Physiol.* **195**, 105576 (2023).
13. Ceseski, A. R., Godar, A. S., Ohadi, S. & Al-Khatib, K. Target and nontarget mechanisms of AHAS inhibitor cross-resistance patterns in *Cyperus difformis*. *Pestic. Biochem. Physiol.* **193**, 105444 (2023).
14. Porebski, S., Bailey, L. G. & Baum, B. R. Modifcation of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Rep.* **15**, 8–15 (1997).
15. Marcais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
16. Vurture, G. W. *et al*. GenomeScope: fast reference-free genome profling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
17. Cheng, H. Y. *et al*. Haplotype-resolved assembly of diploid genomes without parental data. *Nat. Biotechnol.* **40**, 1332–1335 (2022).
18. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
19. Wingett, S. *et al*. HiCUP: pipeline for mapping and processing Hi-C data [version 1; peer review: 2 approved, 1 approved with reservations]. *F1000Research* **4**, 1310 (2015).
20. Burton, J. N. *et al*. Chromosome-scale scafolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**, 1119–1125 (2013).
21. Durand, N. C. *et al*. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
22. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
23. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 1–27 (2020).
24. Flynn, J. M. *et al*. RepeatModeler2 for automated genomic discovery of transposable element families. *PNAS* **117**, 9451–9457 (2020).
25. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21**, i351–i358 (2005).
26. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).
27. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **25**, 4.10.11–14.10.14 (2009).
28. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
29. Kim, D. *et al*. TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
30. Grabherr, M. G. *et al*. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
31. Haas, B. J. *et al*. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).
32. Altschul, S. F. *et al*. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
33. Yu, X. J., Zheng, H. K., Wang, J., Wang, W. & Su, B. Detecting lineage-specific adaptive evolution of brain-expressed genes in human using rhesus macaque as outgroup. *Genomics* **88**, 745–751 (2006).
34. Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. *Genome Res.* **14**, 988–995 (2004).
35. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).
36. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94 (1997).
37. Delcher, A. L., Bratke, K. A., Powers, E. C. & Salzberg, S. L. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**, 673–679 (2007).
38. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
39. Parra, G., Blanco, E. & Guigó, R. GeneID in Drosophila. *Genome Res.* **10**, 511–515 (2000).
40. Haas, B. J. *et al*. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
41. Buchfnk, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
42. Blum, M. *et al*. Te InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* **49**, D344–D354 (2021).
43. Lagesen, K. *et al*. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**, 3100–3108 (2007).
44. Chan, P., Lin, B., Mak, A. J. & Lowe, T. M. tRNAscan-SE 2.0: improved detection and functional classifcation of transfer RNA genes. *Nucleic Acids Res.* **49**, 9077–9096 (2021).
45. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).
46. LI, J. Genome data of *Cyperus difformis*. *GSA. Dataset* https://ngdc.cncb.ac.cn/gsa/browse/CRA031324 (2025).
47. LI, J. & Wang, Y. Genome assembly and annotation file of *Cyperus difformis* L. *FigsharesDataset* https://doi.org/10.6084/m9.figshare.29072027 (2025).
48. LI, J. Cyperus difformis isolate JL-2025, whole genome shotgun sequencing project. NCBI. Dataset. https://www.ncbi.nlm.nih.gov/nuccore/JBTANI000000000 (2025).

Acknowledgements
This study was supported by National Natural Science Foundation of China (32360681 and 32460693), Guangxi Natural Science Foundation (2021GXNSFDA220007 and 2024GXNSFAA010013), Research Funding of Guangxi Academy of Agriculture Sciences (2021YT066, 2022JM43 and 2024ZX17), Guangxi Key Laboratory of Biology for Crop Diseases and Insect Pests (22-035-31-23ST02), Research Funding of Guangxi Vocational University of Agriculture (XKJ2305), and Guangxi Innovation Group of National Modern Agricultural Industry Technology System (nycytxgxcxtd-2024-16-03).

Author contributions
Y.W. and J.L. designed this research. J.Z., W.Z. and Y.M. analyzed the data. M.Q. and W.L. collected the samples. Y.W. and J.L. drafted and revised the manuscript. All co-authors contributed to and approved this manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to Y.W.