scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Chromosome-level genome assembly of narrow-leaf bur-reed (*Sparganium angustifolium* Michx., Typhaceae)

Xiang Shi[1], Jianhua Xue[2,3] & Xinwei Xu[1] ✉

The narrow-leaf bur-reed (*Sparganium angustifolium* Michx.), a member of the Typhaceae family, is an ecologically important species in temperate Northern Hemisphere aquatic ecosystems. As a phylogenetically basal taxon within the order Poales, *S. angustifolium* is critical for understanding evolutionary adaptations to aquatic environments. While molecular mechanisms of aquatic adaptation have been extensively studied in Alismatales, genomic resources for basal Poales remain scarce. Here, we present the first chromosome-level genome assembly of *S. angustifolium*, comprising 486.52 Mb with 99.94% coverage across 15 chromosomes. The assembly demonstrates high continuity, with both contig and scaffold N50 values reaching 33.25 Mb. Genome annotation identified 23,767 protein-coding genes and revealed that repetitive sequences represent 70.16% of the genome. Our findings provide valuable genomic resources for comparative studies of plant adaptation to aquatic environments across different evolutionary lineages.

Background & Summary

Narrow-leaf bur-reed (*Sparganium angustifolium* Michx.), a member of the Typhaceae family, is an ecologically significant species in temperate aquatic ecosystems across the Northern Hemisphere[1,2]. It occupies diverse habitats including lakes, pools, ditches and streams. This species often forms dense local clones with floating leaves occupying substantial portions of the water surface, spreading clonally via stolons. The plant provides shelter for aquatic animals and its fruits and leaves constitute a valuable food resource for wildfowl[2]. *S. angustifolium* is characterized by its slender growth form, typically bearing submerged or floating vegetative structures[2]. Consequently, its genomic resources are vital for exploring the molecular mechanisms underlying plant aquatic adaptation. Current research on such mechanisms has primarily focused on the order Alismatales[3,4]. As a phylogenetically basal lineage within the order Poales[5], *S. angustifolium* represents a critical system for understanding the evolutionary adaptation to aquatic environments. Most species in the Typhaceae are emergent, and several genomes of such species, *S. stoloniferum*[6], *Typha latifolia*[7], and *T. angustifolia*[8], have been sequenced. Unlike these emergent species, which grow along water margins with most of their vegetative structures above water, the floating-leaved *S. angustifolium* inhabits deeper waters, with the majority of its vegetative biomass submerged, thus necessitating distinct aquatic adaptive strategies. Therefore, assembling a high-quality genome of *S. angustifolium* will provide essential genomic resources for elucidating the mechanism of aquatic adaptation within this key family and across the Poales.

Here we integrated Pacific Biosciences (PacBio) high-fidelity (HiFi) sequencing and high-throughput chromosome conformation capture (Hi-C) technology to assemble a chromosome-level genome of *S. angustifolium*. The assemble genome size is 486.52 Mb, with a contig N50 length of 33.25 Mb. We predicted 23,767 protein-coding genes and identified repetitive sequences representing 70.16% of the genome (Table 1). This high-quality genome serves as a key resource for investigating genome evolution and exploring aquatic adaptation mechanism within the Poales.

[1]National Field Station of Freshwater Ecosystem of Liangzi Lake, College of Life Sciences, Wuhan University, Wuhan, 430072, China. [2]Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing, 100093, China. [3]China National Botanical Garden, Beijing, 100093, China. ✉e-mail: xuxw@whu.edu.cn

Genomic features	Statistics
Genome size (Mb)	486.52
Number of chromosomes	15
Number of scaffolds	18
Scaffold N50 size (Mb)	33.25
Chromosomes percent (%)	99.94
Number of contigs	20
Contig Max (Mb)	47.8
Contig N50 size (Mb)	33.25
GC content (%)	42.43
Repetitive sequences (%)	70.16
Number of protein-coding genes	23,767
Average gene length (bp)	5,656
Average CDS length (bp)	1,440
Average exon length (bp)	306

Table 1. Features of the genome assembly of *Sparganium angustifolium*.

Methods

Sample collection and genome sequencing. The sequencing sample of *S. angustifolium* was collected from a river near Genhe City, Inner Mongolia, China (50.7712°N, 121.5027°E). A specimen was deposited at the herbarium of Wuhan University (contact person: Xinwei Xu, xuxw@whu.edu.cn) under the voucher number NMG20240805001. Genomic DNA was extracted using the DNAsecure Plant Kit (Cat. No. DP320-03, Tiangen Biotech). After verification using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific), a Qubit fluorometer (Invitrogen), and gel electrophoresis, the high-quality genomic DNA was used for library construction. The next-generation sequencing (NGS) library was constructed using the NEBNext Ultra II DNA Library Prep Kit according to the manufacturer's instructions. The Hi-C library was prepared from cross-linked chromatins using Belton *et al.*'s protocol[9]. Both the NGS and Hi-C libraries were sequenced on the Illumina NovaSeq 6000 platform to generate 150 bp paired-end reads. Raw reads were subjected to rigorous quality control with the following criteria: (1) reads containing more than 3 ambiguous bases (Ns) in either read of a pair were discarded; (2) read pairs were removed if either read had $\geq$ 20% of bases with a quality score $\leq$ 5; and (3) adapter-contaminated read pairs were excluded. After filtering, a total of 31.68 Gb of NGS data and 90.56 Gb of Hi-C data were retained. In addition, TruSeq mRNA-seq libraries were prepared from root, stem, and leaf tissues and sequenced via NGS, yielding 8.65 Gb, 8.51 Gb, and 8.09 Gb of RNA-seq data, respectively. For PacBio HiFi sequencing, a standard SMRT bell library was constructed using the SMRTbell Express Template Prep Kit 2.0 according to the manufacturer's instructions. Sequencing was then performed on the PacBio Sequel IIe platform, generating 53.27 Gb of raw data.

Genome assembly. The genome size and heterozygosity rate of *S. angustifolium* was estimated using K-mer frequency analysis implemented in Jellyfish v1.1.11[10] based on NGS reads, revealing a genome size of 524.25 Mb and a heterozygosity rate of 0.45% (Fig. 1a). HiFi reads were initially assembled using Hifiasm v0.19.5[11,12] with default parameters, yielding 576.30 Mb of contigs with a contig N50 of 30.64 Mb. To eliminate heterozygous fragments, reads were mapped back to the assembly using Minimap2 v2.26[13], followed by Purge_dups v1.2.5[14]. Contigs were filtered based on NT library alignment, and those with average coverage $< 10\times$ and GC content outside the range of 25.00%–54.00% were discarded using BWA-MEM v0.7.17[15]. This purification and filtering process finally yielded a contig N50 of 33.25 Mb. The draft genome was scaffolded using Hi-C data, a technology that improves assembly quality by capturing the three-dimensional spatial proximity information of chromatin[16]. Specifically, HiC-Pro v2.11.4[17] was used for alignment and 3D-DNA v180922[18] for assembly, followed by manual curation in JuiceBox v1.11.08[19] to correct misassemblies. This process resulted in a final genome size of 486.52 Mb, with 99.94% (486.23 Mb) anchored onto 15 chromosomes and a scaffold N50 of 33.25 Mb (Table 1). To enhance the presentation of the *S. angustifolium* genome assembly, we generated a comprehensive circular genome map using Circos v0.69[20] (Fig. 1b). This visualization provides an intuitive overview of the genomic architecture, which facilitates further analysis of the *S. angustifolium* genome. The number of contigs and scaffolds contained in each chromosome, along with the chromosome length and the number of gaps, were presented in Table 2 (Table 2).

Repeat sequence annotation. We comprehensively characterized repetitive sequences in the *S. angustifolium* genome using both homology-based and de novo approaches. Initial screening with RepeatMasker v4.1.1[21] against the RepBase database[22] was complemented by de novo identification using RepeatModeler v2.0.2a[23], collectively revealing 791,281 repetitive elements spanning 341.38 Mb (70.16% of the genome). Long terminal repeat (LTR) retrotransposons dominated the repeat landscape, occupying 215.56 Mb (44.31% of the genome) and representing 63.1% of all repetitive content (Table 3).

Gene prediction and functional annotation. We performed comprehensive gene annotation through transcriptomic and computational approaches. Total RNA was extracted from root, stem, and leaf tissues of *S. angustifolium* and sequenced as 150 bp paired-end reads on Illumina NovaSeq 6000 platform following library

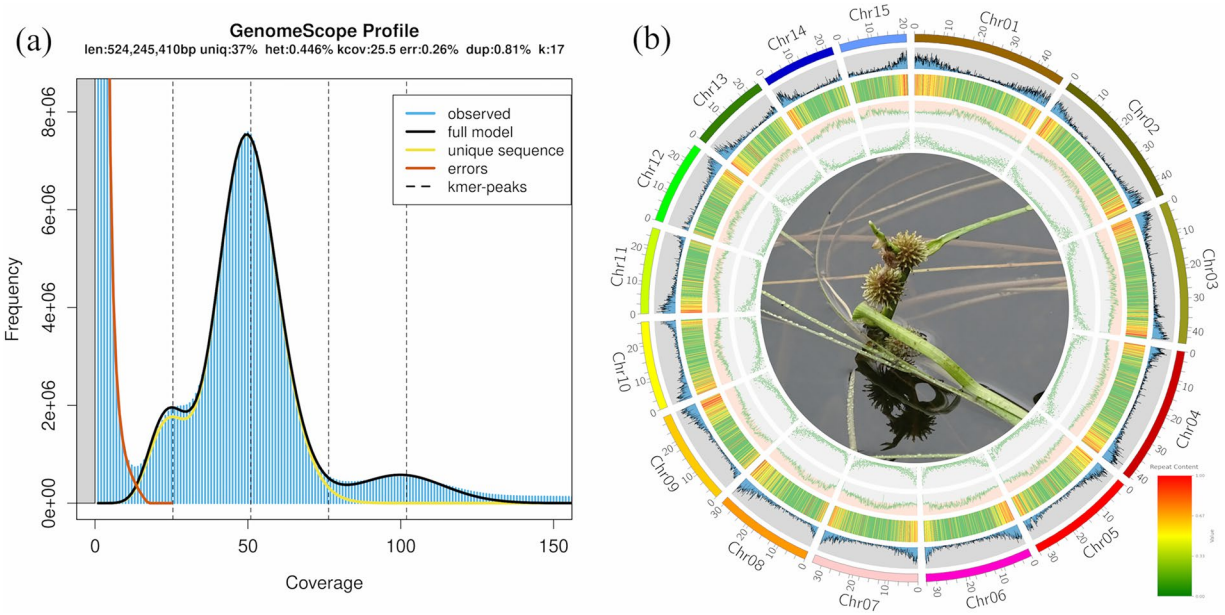

Fig. 1 Overview of the genomic landscape of *Sparganium angustifolium*. (**a**) Genome survey of *S. angustifolium* using k-mer frequency analysis. (**b**) Circle genomic landscape of *S. angstifolium*. The circles from the outside to the inside represent Chromosomes length, gene count, repeat sequence content, GC content, gene density, and the picture of *S. angustifolium*, respectively.

Chromosome	Contig Number	Scaffold Number	Chromosome Length(bp)	Gap Number
Chr01	1	2	47964942	1
Chr02	1	1	44858736	0
Chr03	1	1	44083491	0
Chr04	1	1	42433007	0
Chr05	1	1	33259253	0
Chr06	1	1	33245195	0
Chr07	1	1	32931560	0
Chr08	1	1	30635978	0
Chr09	1	1	29220466	0
Chr10	2	1	27476996	0
Chr11	1	1	26570386	0
Chr12	1	2	25935100	1
Chr13	1	1	24512531	0
Chr14	1	1	22703289	0
Chr15	2	1	20665378	0

Table 2. Chromosome characteristics of Sparganium angustifolium.

preparation and quality control. RNA-Seq reads were aligned to the genome using HISAT2 v2.2.1[24], with transcript assembly performed by Cufflinks v2.2.1[25,26]. Gene prediction combined multiple approaches: de novo prediction using AUGUSTUS v3.5.0[27], SNAP v6.0[28], GlimmerHMM v3.0.4[29], and GeneMark-ET v4.32[30]; homology-based prediction with GeMoMa v1.9[31,32] against four Poales reference genomes: *Sparganium stoloniferum*[33], *Typha angustifolia*[34], *Ananas comosus*[35], and *Tillandsia fasciculata*[36]; and UTR/alternative splicing annotation via PASA v2.5.2[37]. This integrated analysis identified 23,767 protein-coding genes with an average length of 5,656 bp. Functional annotation was performed against GO, KEGG, SWISS-PROT, NR, and eggNOG databases using BLAST v2.2.25[38], Interproscan v5.59-91.0[39], eggnog-mapper v2.1.12[40,41], and KOBAS v3.0[42], with an E-value cutoff of $\leq$1e-5. Overall, 22,907 genes (96.38%) were successfully annotated (Table 4), including 6,109 genes annotated by all five databases (Fig. 2). The NR database provided the most unique annotations (1,339 genes, Fig. 2). In addition, we used tRNAscan-SE v2.0.11[43] to predict tRNAs and Infernal v1.1.4[44] to identify other types of non-coding RNAs (ncRNAs) based on the Rfam database. This analysis revealed 3,332 ncRNA genes spanning 725,610 bp of genomic sequence. The ncRNA repertoire was dominated by rRNAs (2,403 rRNAs totaling 634,680 bp), 397 tRNAs, 279 snoRNAs, 172 miRNAs, and 81 snRNAs (Table 5).

Class	Superfamily	Classification	Number	Length (bp)	Percentage in the genome (%)
ClassI (Retrotransposon)	LTR	LTR	190,758	101,166,754	20.79
		LTR/Cassandra	2,002	57,601	0.01
		LTR/Caulimovirus	4	1,398	0.00
		LTR/Copia	44,455	22,620,562	4.65
		LTR/Gypsy	128,802	91,714,751	18.85
	Non-LTR	LINE/L1	41,395	31,071,705	6.39
		LINE/L1-Tx1	107	13,422	0.00
		LINE/RTE-BovB	521	79,823	0.02
		Retroposon/L1-derived	2	142	0.00
		SINE/tRNA-RTE	5	231	0.00
ClassII(DNA_Transposon)	DNA-TE	DNA	86,309	24,675,517	5.07
		DNA/CMC-EnSpm	8,111	700,980	0.14
		DNA/MULE-MuDR	17,016	7,431,559	1.53
		DNA/MuDR	153	123,849	0.03
		DNA/PIF-Harbinger	7,114	4,455,086	0.92
		DNA/TcMar-Stowaway	655	25,716	0.01
		DNA/Zisupton	152	45,109	0.01
		DNA/hAT-Ac	5,865	2,461,585	0.51
		DNA/hAT-Tag1	4,561	2,292,724	0.47
		DNA/hAT-Tip100	2,189	854,600	0.18
	Helitron	RC/Helitron	4,207	882,751	0.18
Others	DNAauto/PILETAA	DNAauto/PILETAA	235	81,891	0.02
	Satellite	Satellite	166	82,174	0.02
	Satellite/centr	Satellite/centr	28	2,186	0.00
	Satellite/subtelo	Satellite/subtelo	1	48	0.00
	Simple_repeat	Simple_repeat	37	1,299	0.00
	Unknown	Unknown	246,431	50,533,737	10.39
Total			791,281	341,377,200	70.16

Table 3. General statistics of repetitive sequences in the *Sparganium angustifolium* genome.

	Database	Number	Percentage (%)
Total	Protein-coding genes	23,767	100
Annotated	NR	22,869	96.22
	SwissProt	18,050	75.95
	KEGG	9,989	42.03
	GO	10,665	44.87
	EggNOG	21,376	89.94
Not Annotated		860	3.62
Total annotated		22,907	96.38

Table 4. The annotated genes of the *Sparganium angustifolium* genome by different databases.

Data Records

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive in National Genomics Data Center (NGDC), China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences. The accession number of these data is CRA027021[45] under BioSample SAM143465 and BioProject PRJCA041909. The survey reads (accession numbers ERR15844965[46] and ERR15844971[47]), HiFi reads (accession numbers ERR15838631[48], ERR15838536[49], and ERR15838531[50]), and Hi-C reads (accession numbers ERR15842105[51] and ERR15844973[52]) have been deposited in the European Nucleotide Archive. The genome assembly data have been deposited in the Genome Warehouse (GWH) of the National Genomics Data Center (NGDC) with the accession number GWHGEEK00000000.1[53] and in the European Nucleotide Archive with the accession number GCA_977063505[54]. The annotation files of genome assembly are available on the Figshare database[55].

Technical Validation

Following genome assembly, next-generation sequencing data were aligned to the final assembly using BWA-MEM v0.7.17 with default parameters, yielding an alignment rate of 91.19%. This high alignment rate indicates excellent genome assembly completeness. Subsequently, genome coverage was calculated using mosdepth v0.3.1[56], revealing an extremely high coverage rate of 99.97%, further supporting the accuracy of genome assembly.

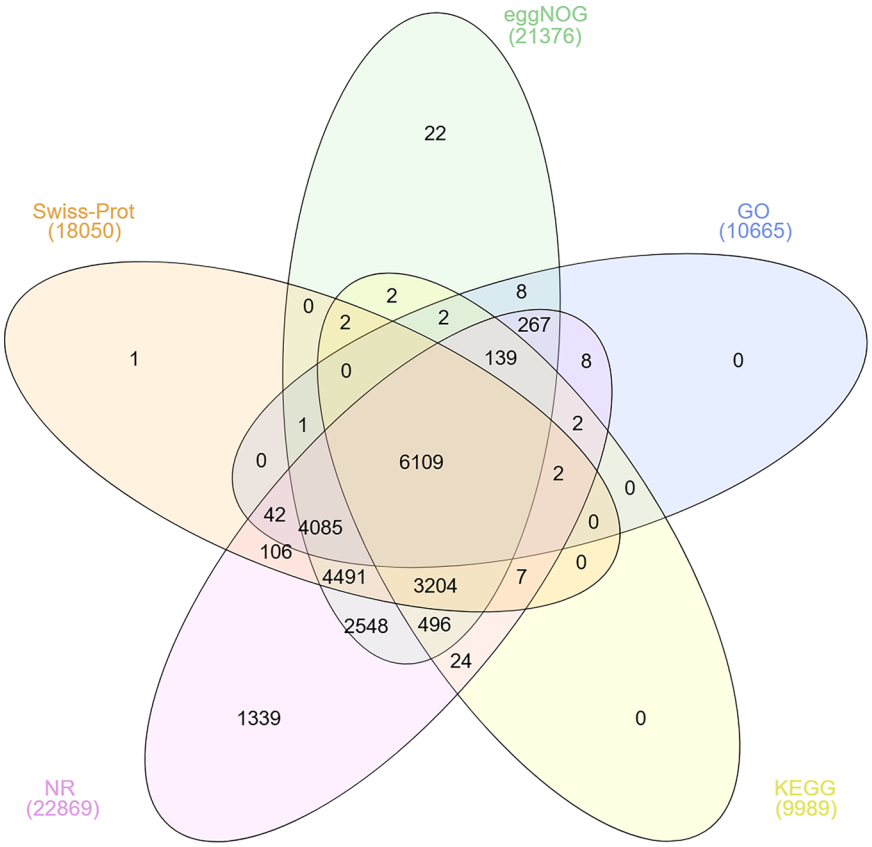

Fig. 2 Venn diagram of the number of genes annotated in different public databases.

Class	Number	Total length (bp)
rRNA	2,403	634,680
miRNA	172	20,215
tRNA	397	29,734
snRNA	81	11,940
snoRNA	279	29,041
Total	3,332	725,610

Table 5. Number and length of predicted non-coding RNAs.

The resulting chromosome-level genome assembly of *Sparganium angustifolium* spans 486.52 Mb, with both contig N50 and scaffold N50 values at 33.25 Mb. The assembled genome size is notably less than the K-mer estimate (524.25 Mb), a discrepancy likely attributable to heterozygosity or repetitive sequences inflating the k-mer prediction. Assessment of genome completeness using the Liliopsida_odb12 database in BUSCO v5.8.3 showed BUSCO completeness scores of 96.4% (genome) and 97.0% (protein) (Fig. 3a). To further validate assembly accuracy, Hi-C interaction matrices were generated using Samblaster v0.1.26[57] and HapHic v1.0.7[58]. The resulting heatmap displayed distinct clustering pattern, with strong interactions between adjacent regions and weaker interactions elsewhere (Fig. 3b), indicative of high-quality chromosome-level assembly.

Data availability

The raw sequence data have been deposited in the Genome Sequence Archive at the National Genomics Data Center under accession number CRA027021[45]. The survey, HiFi, and Hi-C reads have been deposited in the European Nucleotide Archive under accession numbers ERR15844965[46] and ERR15844971[47] (survey reads); ERR15838631[48], ERR15838536[49], and ERR15838531[50] (HiFi reads); and ERR15842105[51] and ERR15844973[52] (Hi-C reads). The genome assembly has been deposited in the Genome Warehouse under accession number GWHGEEK00000000.1[53] and in the European Nucleotide Archive under accession number GCA_977063505[54]. The genome annotation files are available in the Figshare database(https://doi.org/10.6084/m9.figshare.29375852)[55].

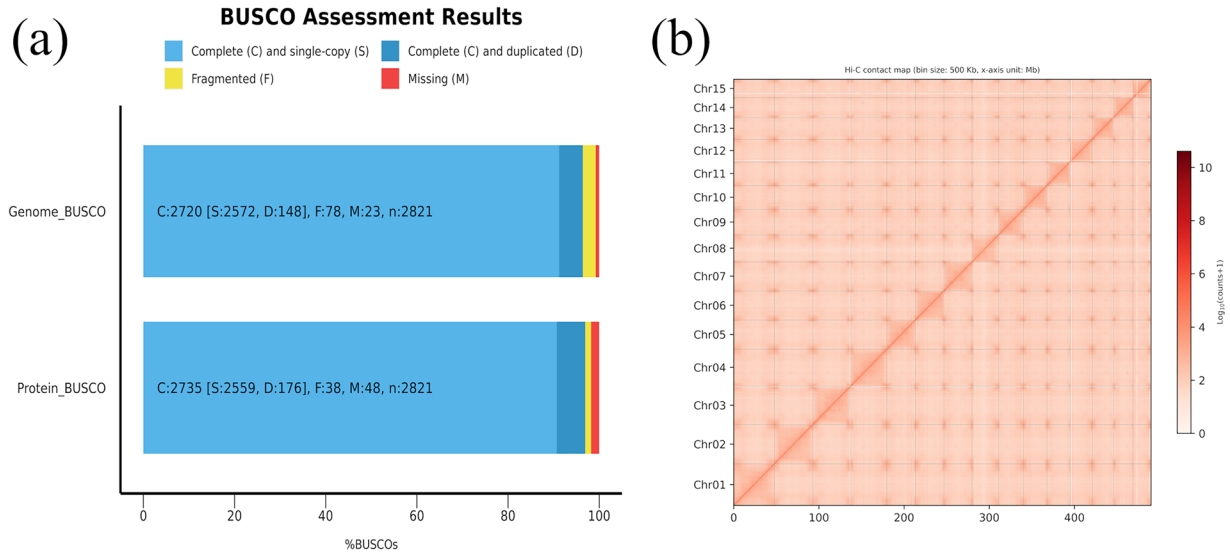

Fig. 3 Genome assembly quality assessment. (**a**) BUSCO assessment of genetic integrity of *Sparganium angustifolium*. (**b**) Hi-C interaction heat map of *S. angustifolium*.

Code availability

All bioinformatics tools and software used for genome assembly, annotation, and data analysis in this study were operated strictly according to their official user manuals, with no custom code employed. Software versions and parameters are comprehensively documented in the Methods section.

Received: 4 July 2025; Accepted: 15 January 2026;
Published online: 21 January 2026

References

1. The Angiosperm Phylogeny Group. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* **161**, 105–121, https://doi.org/10.1111/j.1095-8339.2009.00996.x (2009).
2. Cook, C. D. K. & Nicholls, M. S. A monographic study of the genus *Sparganium*. Part 1: Subgenus *Xanthosparganium*. *Bot. Helv.* **96**, 213–267, https://doi.org/10.5169/seals-67202 (1986).
3. Chen, L. Y. *et al.* Phylogenomic analyses of Alismatales shed light into adaptations to aquatic environments. *Mol. Biol. Evol.* **39**, msac079, https://doi.org/10.1093/molbev/msac079 (2022).
4. Ma, X. *et al.* Seagrass genomes reveal ancient polyploidy and adaptations to the marine environment. *Nat. Plants.* **10**, 240–255, https://doi.org/10.1038/s41477-023-01608-5 (2024).
5. Givnish, T. J. *et al.* Assembling the tree of the monocotyledons: Plastome sequence phylogeny and evolution of Poales. *Ann. Mo. Bot. Gard.* **97**, 584–616, https://doi.org/10.3417/2010023 (2010).
6. Zou, Y. *et al.* Genomic analysis of the emergent aquatic plant *Sparganium stoloniferum* provides insights into its clonality, local adaptation and demographic history. *Mol. Ecol. Res.* **23**, 1868–1879, https://doi.org/10.1111/1755-0998.13850 (2023).
7. Shane, D. W., Joanna, R. F., Xu, X. W. & Aaron, B. A. S. Genome assembly, annotation, and comparative analysis of the cattail Typha latifolia. *G3-Genes Genom. Genet.* **12**, jkab401, https://doi.org/10.1093/g3journal/jkab401 (2022).
8. Liao, Y. *et al.* Chromosome-level genome and high nitrogen stress response of the widespread and ecologically important wetland plant *Typha angustifolia*. *Front. Plant Sci.* **14**, 1138498, https://doi.org/10.3389/fpls.2023.1138498 (2023).
9. Belton, J. M. *et al.* Hi - C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276, https://doi.org/10.1016/j.ymeth.2012.05.001 (2012).
10. Marcais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics.* **27**, 764–770, https://doi.org/10.1093/bioinformatics/btr011 (2011).
11. Liu, B. *et al.* Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. Preprint at https://arxiv.org/abs/1308.2012 (2013).
12. Cheng, H. *et al.* Haplotype-resolved assembly of diploid genomes without parental data. *Nat. Biotechnol.* **40**, 1332–1335, https://doi.org/10.1038/s41587-022-01261-x (2022).
13. Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics.* **37**, 4572–4574, https://doi.org/10.1093/bioinformatics/btab705 (2021).
14. Guan, D. *et al.* Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics.* **36**, 2896–2898, https://doi.org/10.1093/bioinformatics/btaa025 (2020).
15. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://arxiv.org/abs/1303.3997 (2013).
16. Zhu, K. *et al.* Advancing chromosomal-scale, haplotype-resolved genome assembly: beading with Hi-C data. *Adv. Biotechnol.* **2**, 28, https://doi.org/10.1007/s44307-024-00035-7 (2024).
17. Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259, https://doi.org/10.1186/s13059-015-0831-x (2015).
18. Dudchenko, O. *et al.* De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science.* **356**, 92–95, https://doi.org/10.1126/science.aal3327 (2017).
19. Neva, C. D. *et al.* Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101, https://doi.org/10.1016/j.cels.2015.07.012 (2016).
20. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645, https://doi.org/10.1101/gr.092759.109 (2009).

21. Saha, S., Bridges, S., Magbanua, Z. V. & Peterson, D. G. Empirical comparison of ab initio repeat finding programs. *Nucleic Acids Res.* **36**, 2284–2294, https://doi.org/10.1093/nar/gkn064 (2008).
22. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. Dna.* **6**, 1–6, https://doi.org/10.1186/s13100-015-0041-9 (2015).
23. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580, https://doi.org/10.1093/nar/27.2.573 (1999).
24. Kim, D. *et al*. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915, https://doi.org/10.1038/s41587-019-0201-4 (2019).
25. Trapnell, C. *et al*. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515, https://doi.org/10.1038/nbt.1621 (2010).
26. Trapnell, C. *et al*. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31**, 46–53, https://doi.org/10.1038/nbt.2450 (2013).
27. 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, https://doi.org/10.1093/bioinformatics/btn013 (2008).
28. Johnson, A. D. *et al*. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics.* **24**, 2938–2939, https://doi.org/10.1093/bioinformatics/btn564 (2008).
29. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics.* **20**, 2878–2879, https://doi.org/10.1093/bioinformatics/bth315 (2004).
30. Lomsadze, A., Burns, P. D. & Borodovsky, M. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res.* **42**, e119, https://doi.org/10.1093/nar/gku557 (2014).
31. Keilwagen, J. *et al*. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res.* **44**, e89, https://doi.org/10.1093/nar/gkw092 (2016).
32. Keilwagen, J. *et al*. Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinformatics.* **19**, 1–12, https://doi.org/10.1186/s12859-018-2203-5 (2018).
33. Zou, Y. The chromosome-level genome of *Sparganium stoloniferum*. *figshare. Dataset.* https://doi.org/10.6084/m9.figshare.21407208.v1 (2022).
34. Chen, L. Dataset: Genome assemblies of *Typha angustifolia*. *figshare. Dataset.* https://doi.org/10.6084/m9.figshare.25769703.v3 (2024).
35. *NCBI GenBank* https://identifiers.org/ncbi/insdc:LODP00000000.1 (2016).
36. *NCBI GenBank* https://identifiers.org/ncbi/insdc:JAQOTM000000000.1 (2024).
37. Haas, B. J. *et al*. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666, https://doi.org/10.1093/nar/gkg770 (2003).
38. Altschul, S. F. *et al*. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402, https://doi.org/10.1093/nar/25.17.3389 (1997).
39. Zdobnov, E. M. & Apweiler, R. InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics.* **17**, 847–848, https://doi.org/10.1093/bioinformatics/17.9.847 (2001).
40. Cantalapiedra, C. P. *et al*. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829, https://doi.org/10.1093/molbev/msab293 (2021).
41. 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**, D309–D314, https://doi.org/10.1093/nar/gky1085 (2019).
42. Bu, D. *et al*. KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res.* **49**, W317–W325, https://doi.org/10.1093/nar/gkab447 (2021).
43. Chan, P. P. *et al*. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* **49**, 9077–9096, https://doi.org/10.1093/nar/gkab688 (2021).
44. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics.* **29**, 2933–2935, https://doi.org/10.1093/bioinformatics/btt509 (2013).
45. *NGDC Genome Sequence Archive* https://ngdc.cncb.ac.cn/gsa/browse/CRA027021 (2025).
46. *European Nucleotide Archive* https://identifiers.org/insdc.sra:ERR15844965 (2025).
47. *European Nucleotide Archive* https://identifiers.org/insdc.sra:ERR15844971 (2025).
48. *European Nucleotide Archive* https://identifiers.org/insdc.sra:ERR15838631 (2025).
49. *European Nucleotide Archive* https://identifiers.org/insdc.sra:ERR15838536 (2025).
50. *European Nucleotide Archive* https://identifiers.org/insdc.sra:ERR15838531 (2025).
51. *European Nucleotide Archive* https://identifiers.org/insdc.sra:ERR15842105 (2025).
52. *European Nucleotide Archive* https://identifiers.org/insdc.sra:ERR15844973 (2025).
53. *NGDC Genome Warehouse* https://bigd.big.ac.cn/gwh/Assembly/98122/show (2025).
54. *European Nucleotide Archive* https://identifiers.org/insdc.gca:GCA_977063505 (2025).
55. Shi, X., Xue, J. & Xu, X. Chromosome-level genome assembly of narrow-leaf bur-reed (Sparganium angustifolium Michx., Typhaceae). *figshare. Dataset.* https://doi.org/10.6084/m9.figshare.29375852 (2025).
56. Pedersen, B. S. & Quinlan, A. R. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics.* **34**, 867–868, https://doi.org/10.1093/bioinformatics/btx699 (2018).
57. Faust, G. G. & Hall, I. M. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics.* **30**, 2503–2505, https://doi.org/10.1093/bioinformatics/btu314 (2014).
58. Zeng, X. *et al*. Chromosome-level scaffolding of haplotype-resolved assemblies using Hi-C data without reference genomes. *Nat. Plants.* **10**, 1184–1200, https://doi.org/10.1038/s41477-024-01755-3 (2024).

Acknowledgements

This study was supported by the National Water Pollution Control and Treatment Science and Technology Major Project, China (No. 2015ZX07503005). The calculations in this paper were performed using the supercomputing system at the Supercomputing Center of Wuhan University.

Author contributions

X.X. designed the research; J.X. carried out the field collections; X.S. carried out the experiments and performed the data analysis; X.S., J.X. and X.X. wrote and revised the manuscript. All authors read and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to X.X.

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.