the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2025) **124,** e70558 doi: 10.1111/tpj.70558

RESOURCE

The T2T genome assembly of watershield (*Brasenia schreberi*) unveils genomic insights into aquatic adaptation

Huiying Shang[1,†], Changhong Luo[2,†] and Yongzhi Yang[2,*]

[1]*Xi'an Botanical Garden of Shaanxi Province (Institute of Botany of Shaanxi Province), Xi'an, Shaanxi Province, China, and* [2]*State Key Laboratory of Herbage Improvement and Grassland Agro-Ecosystems, College of Ecology, Lanzhou University, Lanzhou 730000, China*

Received 27 June 2025; revised 1 September 2025; accepted 21 October 2025.
*For correspondence (e-mail yangyz@lzu.edu.cn).
†These authors contributed equally to this work.

SUMMARY

Watershield (*Brasenia schreberi*), belonging to Cabombaceae within the order Nymphaeales, represents one of the early-diverged angiosperm lineages. This perennial floating leaf freshwater aquatic plant features submerged juvenile leaves enveloped in a thick layer of transparent gelatinous mucilage, aiding in its resistance to aquatic stress. However, the evolutionary history of the mechanisms underlying its specific phenotype remains unclear. In this study, we present the telomere-to-telomere level genome of *B. schreberi*, unveiling that it underwent two rounds of whole-genome duplications (WGDs) and a recent whole-genome triplication, with the most ancient WGD being shared by Nymphaeaceae. WGD and dispersed duplication significantly contributed to the expansion of gene families, which are primarily associated with environmental adaptation. Additionally, we discovered that mature leaves primarily conduct photosynthesis and may transport nutrients to underwater juvenile leaves for polysaccharide synthesis. We also identified an ancestral broad expression pattern of ABC genes, and the similar expression of anthocyanin biosynthesis genes across all flower organs resulted in entirely purple flowers. Our findings deepen the understanding of the evolution of this specific aquatic plant phenotypes.

Keywords: T2T genome, *Brasenia schreberi*, Nymphaeales evolution, polysaccharide synthesis.

INTRODUCTION

Through evolutionary time, plants have developed intricate adaptive strategies through dynamic interactions with their environments, evolving specialized physiological and morphological traits to cope with environmental variability (Bowles et al., 2023). Aquatic plants in particular face unique challenges, as aquatic ecosystems present fundamentally distinct physicochemical conditions from terrestrial habitats, including oxygen limitation, light attenuation, hydrodynamic forces, and thermal fluctuations (Anderson et al., 2011; Titus & Urban, 2009), driving the evolution of sophisticated adaptive traits in these organisms (Li et al., 2024). A prime example is *Brasenia schreberi* ($2n = 72$), a perennial aquatic angiosperm of the Cabombaceae family (Nymphaeales) distributed across Asia (Lu et al., 2023; Zhang, Chen, et al., 2020; Zhang, Zhang, et al., 2020), North America, and Oceania. This species exhibits remarkable adaptations including submerged stem apices and developing leaves enveloped in a transparent mucilaginous matrix that dissipates during leaf maturation and emergence. The edible juvenile shoots and developing leaves, characterized by their thick mucilage coating, have been traditionally utilized in Chinese cuisine and medicine (Ai et al., 2024; Hongyi et al., 2024). Biochemical analyses reveal this mucilage contains valuable polysaccharides, proteins, vitamins, and bioactive compounds with demonstrated antimicrobial, anti-inflammatory, and hypoglycemic properties (Wang et al., 2023; Xiao et al., 2016). However, the photoassimilates translocation and thick mucilage biosynthesis remain to be elucidated.

While recent transcriptomic, chloroplast genomic and whole-genome studies have firmly established *B. schreberi's* taxonomic position within Cabombaceae as sister to Nymphaeaceae (Lu et al., 2023; Zhang, Chen, et al., 2020; Zhang, Zhang, et al., 2020), its genomic

evolutionary trajectory remains incompletely understood. Despite synonymous site divergence (*Ks*) analyses, whether *B. schreberi* shares its whole-genome duplication (WGD) event with other Nymphaeaceae species remains unresolved (Wen et al., 2025; Zhang, Chen, et al., 2020; Zhang, Zhang, et al., 2020), necessitating integrated methods for clarification. Moreover, angiosperms exhibit remarkable plasticity in floral pigmentation, showing substantial interspecific and intraspecific variation (Amborella Genome Project et al., 2013; Gigord et al., 2001; Hopkins, 2013; Wessinger et al., 2012). In striking contrast to the showy, multicolored blooms of water lilies in the closely related Nymphaeaceae family, *B. schreberi* displays small, uniformly purplish-red flowers with undifferentiated perianth parts; yet its evolutionary mechanism remains unclear.

In this study, we successfully generated a high-quality telomere-to-telomere (T2T) genome assembly for *B. schreberi* through integrated sequencing approaches combining PacBio HiFi, ONT ultra-long reads, and Hi-C data. Our comprehensive genomic analyses identified two WGD events and one whole-genome triplication (WGT), with comparative genomic evidence confirming the most ancient WGD as a shared evolutionary event across Nymphaeaceae species. By employing multi-omics strategies, we systematically reconstructed metabolic pathways for flavonoids and anthocyanins while elucidating differential gene expression patterns associated with mucilage distribution, a key adaptive feature in this aquatic angiosperm. These findings significantly advance our understanding of genome evolution and phenotypic adaptation in this early-diverging flowering plant lineage.

RESULTS AND DISCUSSION

Implementation of T2T assembly and annotation

Genome analysis of *B. schreberi* using Illumina short-read sequencing estimated a genome size of 1048.18 Mb with 0.621% heterozygosity (Figure S1). For high-quality genome assembly, we generated 87.79 Gb of PacBio HiFi reads (~79.81× coverage) supplemented with ~9.09× coverage of Oxford Nanopore (ONT) ultra-long reads to improve assembly continuity (Table S1). The assembly pipeline integrated HiFi, Hi-C, and ONT data using hifiasm, producing contigs totaling 1.10 Gb with an N50 of 26.83 Mb (Table 1; Table S4). Using Juicer, we aligned 115.17 Gb of Hi-C data to these contigs and scaffolded them onto 36 chromosomes with 3D-DNA, achieving 96.91% anchoring efficiency (Figure 1b). Subsequent gap-filling with HiFi and ONT reads yielded a final assembly containing only four remaining gaps, including 33 completely gapless chromosomes. Our assembly successfully identified all 36 centromeres and 72 telomeres, confirming T2T status for all chromosomes, with 33 reaching complete T2T gapless level (Figure 1a; Table 1). Notably, *B. schreberi* exhibits unique (AAAGGATAAGAC)n telomeric repeats, differing from the canonical (TTTAGGG)n sequence found in most angiosperms (Table S3; Fajkus et al., 2005).

The quality of our T2T genome assembly was rigorously assessed through multiple evaluation metrics. Comparative analysis with three previous assemblies demonstrated significant improvements in contiguity, with the contig N50 increasing from 7.1 (V1) to 26.7 Mb (V4_T2T) and a substantial reduction in gap regions (Table 1; Figure 1d). Comprehensive quality assessments

Table 1 Comparison of four *Brasenia schreberi* assemblies

	B. schreberi			
Type	V1	V2	V3	V4_T2T
Assembly				
Assembly level	Chromosome	Contig	Chromosome	Chromosome
Total length of Contigs (bp)	1 170 389 352	1 055 148 839	1 011 141 865	1 088 856 280
Contig N50 (bp)	7 141 099	22 379 495	49 104	26 836 987
Gaps	1674	—	139 348	4
Telomeres	—	—	—	72
LTR Assembly Index, LAI	6.21	5.99	4.89	8.48
BUSCO (embryophyta)	94.3	95.4	90.8	96.4
Annotation				
Percentage of repeat elements	47.69	49.42	48.73	49.89
Protein-coding genes	74 699	67 747	74 426	70 821
Average gene length (bp)	6053.912	5421.904	4549.780	5108.817
Average CDS length (bp)	1018.713	952.981	934.155	1159.819
Average exons per gene (bp)	4.155	4.092	4.055	4.407
Average exon length (bp)	245.177	232.871	219.092	263.134
Average Intron Length (bp)	1356.679	1261.722	1198.541	1158.844
Reference	Lu et al. (2023)	Wu et al. (2024)	Wen et al. (2025)	

further validated the assembly's robustness, including superior Long Terminal Repeats Assembly Index (LAI) scores compared to earlier versions. The assembly achieved a BUSCO completeness of 96.4%, surpassing most published Nymphaea genomes (Povilus et al., 2020; Zhang, Chen, et al., 2020; Zhang, Zhang, et al., 2020), along with a consensus quality value of 41.7 (Figure S5), meeting reference-grade standards. Synteny analysis with two published chromosome-level genomes (Lu et al., 2023; Wen et al., 2025) revealed strong collinearity without inter-chromosomal rearrangements, consistent with Hi-C clustering patterns across all assemblies (Figures S2 and S3). In particular, the longest chromosome (Chr12) showed excellent linear synteny between the V4_T2T and V1 versions (Figure 1b). However, other chromosomes exhibited numerous intra-chromosomal structural variations, with a typical example being a large-scale intra-chromosomal translocation detected in Chr35 when comparing the assemblies (Figure 1d). We then examined whether this translocation was a misassembly in different versions or a natural variation between different individuals. Hi-C data re-anchoring confirmed this as a misassembly in the V1 version, whereas our T2T assembly exhibited stronger intra-chromosomal interaction signals and proper contig ordering (Figure 1b,c; Figure S4). These findings collectively demonstrate the superior accuracy and reliability of our T2T-level *B. schreberi* genome assembly.

Our comprehensive analysis of repetitive elements in the *B. schreberi* genome, employing both ab initio prediction and homology-based approaches, revealed 543.34 Mb of repetitive sequences representing 49.33% of the genome. Among these, Long Terminal Repeats (LTRs) constituted the predominant class, accounting for 33.28% of the genomic content, with Gypsy (14.55%) and Copia (10.87%) elements emerging as the most abundant LTR subfamilies (Table S4). Gene prediction analysis identified 70 821 protein-coding genes exhibiting an average coding sequence length of 2659.37 bp, containing exons averaging 263.13 bp in length and introns averaging 1158.84 bp (Table S4). Quality assessment through OMArk and BUSCO analyses demonstrated exceptional gene set completeness, with scores of 98.00 and 91.20%, respectively (Figure S5; Table S4). Functional annotation further established that 84.76% of predicted genes share significant homology with documented proteins in the NCBI-NR, Gene Ontology (GO), and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases (Table S5).

Polyploidization histories of Nymphaeales

Through a multi-method approach, we systematically investigated the polyploidization history of Nymphaeales species. Analysis of synonymous substitution rates (*Ks*) distribution using syntenic paralogs revealed three polyploidization events in *B. schreberi* (*Ks* peaks: 0.11, 0.62, and 1.11), along with one event in *Nymphaea* (*Ks* peak: 0.98) and two in *Euryale* (*Ks* peaks: 0.18 and 1.09), consistent with previous studies (Figure 2b) (Zhang, Chen, et al., 2020; Zhang, Zhang, et al., 2020). Syntenic dot plot analysis identified two intact chromosome-level synteny blocks in *B. schreberi* ($Ks \approx 0.11$), indicating a recent WGT event, while the presence of triplicated ($Ks = 0.62$) and hexaploid ($Ks = 1.11$) (Figure 2b) synteny blocks confirmed two ancient WGD events. Similarly, *Euryale ferox* exhibited one ancient WGD and one recent WGT, whereas *Nymphaea colorata* showed only one WGD event. These findings were further supported by interspecific syntenic depth ratios (*B. schreberi*–*Victoria cruziana* 12:2; *B. schreberi*–*N. colorata* 12:2; *B. schreberi*–*E. ferox* 12:6) and aligned with existing research (Figure 2d) (Wu et al., 2022; Yang et al., 2020).

Further analyses revealed that species from both Nymphaeaceae and Cabombaceae share an ancient WGD event. All three species exhibited similar *Ks* peak characteristics that exceeded their divergence values (Figure 2b). Interspecific synteny plots between *N. colorata* and *B. schreberi* revealed two synteny block groups with different *Ks* values, indicating their shared ancient WGD (Figure 2d). A similar pattern was also observed in the *B. schreberi*–*E. ferox* and *N. colorata*–*E. ferox* comparisons (Figure 2d; Figure S7). Finally, collinear gene tree analysis showed that 91.7% of gene trees supported this shared WGD event in both Nymphaeaceae and Cabombaceae species (Figure 2c).

Phylogenomic and expanded gene family analyses

A total of 1476 single-copy orthologous genes were retrieved by SonicParanoid among nine species, which included one gymnosperm (*Ginkgo biloba*) as the outgroup, four species from the ANA-grade (*Amborella trichopoda*, *E. ferox*, *V. cruziana*, *N. colorata*, and *B. schreberi*), one monocot (*Acorus gramineus*), two eudicots (*Nelumbo nucifera* and *Vitis vinifera*). A highly supported species tree was obtained through maximum-likelihood (ML) analysis of the concatenated nucleotide sequences. The phylogeny tree showed that Amborella and Nymphaeales are successively sister to all other angiosperms (Figure 2a). Within the order Nymphaeales, *Brasenia* diverged from the common ancestor shared with *Euryale*, *Victoria* and *Nymphaea* approximately 117.94 million years ago (Mya), marking a key divergence event within the family during the mid-Cretaceous period (Figure 2a). Interestingly, further resolution within the Nymphaeaceae reveals that the genus *Nymphaea* diverged earlier, around 54.13 Mya, forming a distinct lineage. Subsequently, the genera *Victoria* and *Euryale* form a sister clade that split approximately 33.54 Mya.

Further analysis of gene expansion identified 3841 expanded gene families in the ancestral Nymphaeales

lineage. The GO enrichment analysis revealed a significant enrichment for traits associated with aquatic adaptation (Table S15), such as 'response to hypoxia' and 'response to ethylene', suggesting a common aquatic adaptation in Nymphaeales species. Furthermore, it was determined that 69.98% (34 342 genes) of all expanded genes originated

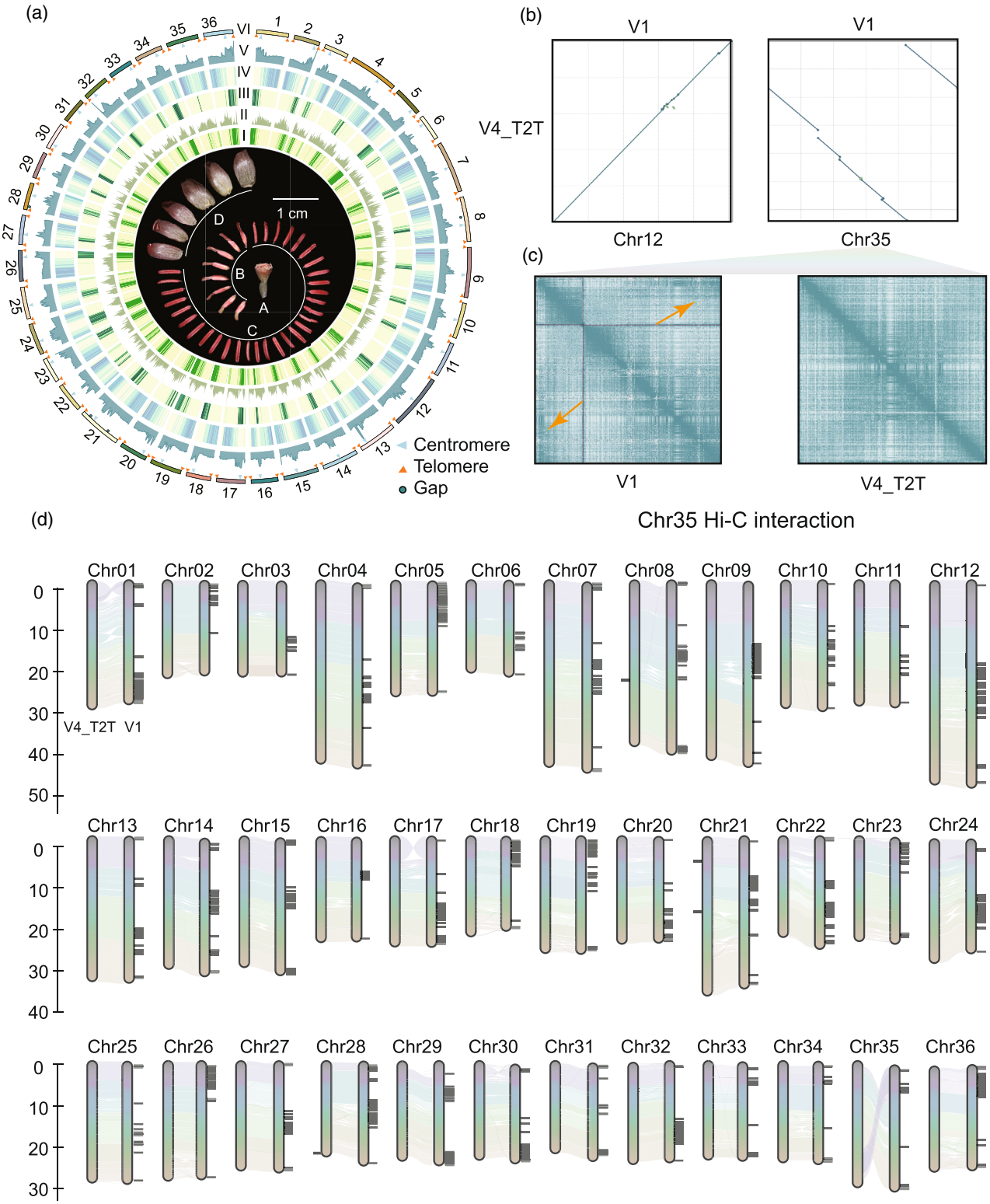

Figure 1. Assembly and annotation of V4_T2T *Brasenia schreberi*.
(a) The center of the circle diagram is the flower organ display of *B. schreberi*. A–D are receptacle, pistil, stamen, petal, sepal. The different trajectories (extending outward) are as follows: (I) Long Terminal Repeats (LTR) density; (II) LTR-*Copia* element density; (III) LTR-*Gypsy* element density; (IV) gene density; (V) GC content; (VI) chromosomes. The values were calculated using a 500 kb sliding window. The telomere and centromere regions are represented by red and blue triangles, respectively. Gray circles indicate chromosomal gaps.
(b) Comparative sequence alignment of Chr12 and Chr35 between V1 and V4_T2T genome assemblies. Connecting lines represent syntenic relationships between homologous chromosomal regions.
(c) Chromosome conformation capture (Hi-C) interaction heatmaps of Chr35 for V4_T2T and V1 genome assemblies, displayed separately.
(d) Genome synteny comparison: V4_T2T, V1 assemblies. Paired homologous chromosomes are shown. V1 and V4_T2T have slim black lines indicating gap positions.

from duplication events (Table S9). Especially, WGD and dispersed duplication (DSD) were identified as the most significant contributors to gene expansion, representing 35.10 and 29.15%, respectively, among all duplication-derived expanded genes (Table S9). These duplication processes are proposed to have played a crucial role in shaping the phenotype and local adaptation of Nymphaeales species. We further focused on the ERF-VII subfamily, which plays a critical role in hypoxia sensing in plants (Sasidharan & Mustroph, 2011) and has been recognized as a central regulator of flooding or hypoxia stress responses in various crops (Wei et al., 2019; Yu et al., 2019). Significantly expanded gene numbers were detected in *B. schreberi* (19 genes) and *E. ferox* (14 genes) compared to *A. trichopoda* (2 genes) and *Arabiopsis thaliana* (5 genes), while *N. colorata* possessed the same gene number as *A. trichopoda*. We also found that WGD is the primary factor contributing to the expansion of this gene family in *B. schreberi* and *E. ferox* (Figure S13), which supports the conclusion that WGD is a major driver of gene family expansion.

Finally, we also identified 7909 gene families that expanded in the *B. schreberi* lineage, and 840 and 5475 expanded gene families in *N. colorata* and *E. ferox*, respectively (Table S8). These expanded genes were mainly related to species-specific local adaptations, such as 'response to freezing' (GO:0050826) and 'response to blue light' (GO:0009637) in *B. schreberi* (Table S6); 'cellular response to salt' (GO:0071366) in *N. colorata* (Table S16); and 'cellular response to cold' (GO:0070417) in *E. ferox* (Table S17). We also noticed an interesting event: the expanded genes of *B. schreberi* also enriched in secondary metabolic processes that may be related to its unique mucilage synthesis ability in Nymphaeales. These enriched GO terms included 'regulation of flavonoid biosynthetic process' (GO:0009962), 'anthocyanin-containing compound biosynthetic process' (GO:0009718), 'antibiotic biosynthetic process' (GO:0017000), and 'terpenoid catabolic process' (GO:0016115) (Table S6).

Flower development-related genes in *B. schreberi*

The *B. schreberi* lineage descended from one of the early-diverged angiosperms, which offers a unique window into the early evolution of the flower organs, especially as it has bisexual flowers, and the petals are very similar to the calyx (Figure 1a; Figure 3a). We mainly focused on the MADS-box genes, which are important regulators of flower development (Ma et al., 2021; Qiu et al., 2024; Zhang et al., 2024). We identified 146 MADS-box genes in *B. schreberi* with phylogenetic distributions across all major lineages identified in angiosperms (Table S10). Among them, 55 genes belonging to type II were further clustered into 15 lineages (Figure S12); these lineages were highly consistent with those genes in *Amborella* and *N. colorata*, which supported all gene lineages of MADS-box formed in the ancestor of angiosperms (Amborella Genome Project et al., 2013; Zhang, Chen, et al., 2020; Zhang, Zhang, et al., 2020). We also found that *FLC* only evolved in the eudicots, and *AGL15* and *OsMADS3252* were specifically lost in the monocots and eudicots, respectively (Arora et al., 2007; Ma et al., 2021) (Table S10). Homologs of the genes for the ABCE model of floral organ identities in *B. schreberi* were also found, including six *AGL6*s (A function for sepals and petals), five *AP3*s and one *PI* (B function for petals and stamen), three *AG*s and three *STK*s (C function for stamen and pistil), and four *SEP* (E function for interacting with ABC function proteins) (Table S10).

To gain more insight into the functions of the MADS-box homologs in *B. schreberi*, we performed RNA-seq for different tissues (Figure S9) and the expression levels of these genes were determined in the four floral organs (sepal, petal, stamen and pistil), leaves, roots and stems. The expression profiles of *B. schreberi* ABCE homologs were mainly expressed in the floral organs and also largely agree with their putative ascribed roles in floral organ patterning (Figure 3b). Two of the six *B. schreberi AGL6* homologs showed a high expression in sepals and petals, consistent with the A-function gene. The other four *AGL6*s showed a low expression in all the floral organs. Four of the five *AP3*s (B-function genes) showed a broad expression in all the four floral organs with the highest expression in stamen, while the last *AP3* showed a very low expression in all floral and non-floral organs. The other B-function gene *PI* showed a high expression in both stamen, petal and sepal. Among the C-function homologs

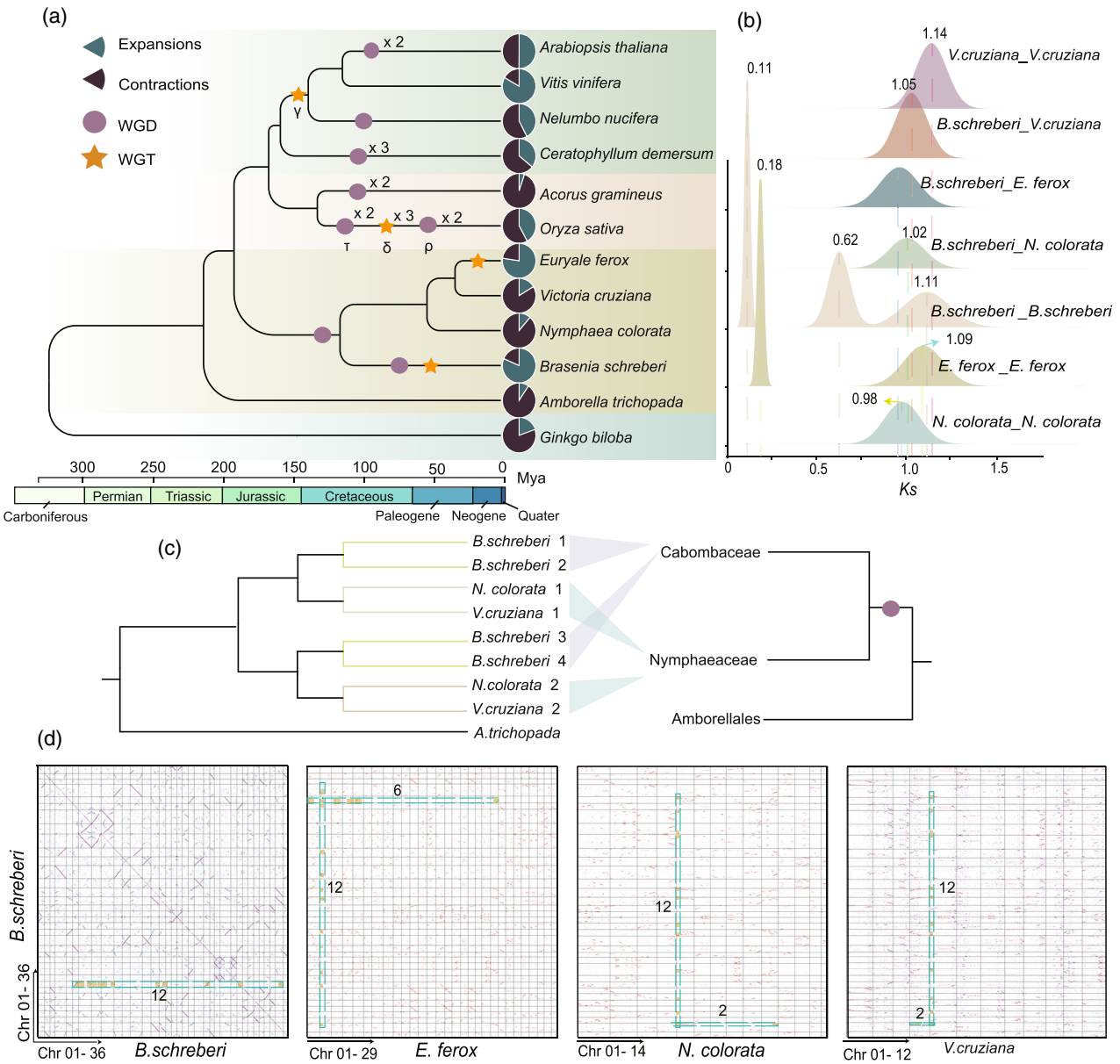

Figure 2. Genome evolution analyses.
(a) Illustration of phylogenetic relationships among nine species. Whole-genome duplication and whole-genome triplication events are represented by purple circles and yellow stars, respectively, at the end of the terminal branches represents the proportion of gene families that underwent expansion (blue) or contraction (black) when comparing with their most recent common ancestor.
(b) Distribution of synonymous substitutions (*Ks*) for collinear gene pairs.
(c) The phylogenetic topology of Nymphaeales constructed from collinear genes extracted from AAK protochromosomes, with Amborella as the outgroup (numerals 1–4 denote subgenomes).
(d) Collinearity dot plots comparing *Brasenia schreberi* with *B. schreberi*, *Euryale ferox*, *Nymphaea colorata* and *Victoria cruziana*.

of *B. schreberi*, three *AG*s are highly expressed in stamen and pistil, while the *STK* gene is mainly highly expressed in stamen. This evidence showed that the ABCE homologs in *B. schreberi* generally exhibit wider ranges of expression in floral organs than their counterparts in eudicot model systems, which may represent the ancestral characters just like those previously reported in *N. colorata*, *Persea americana* (Chanderbali et al., 2008) and *Chloranthus sessilifolius*. Besides the ABCE genes, we also found that *AGL15*s showed a relatively high expression in stamen and one *AGL32* showed higher expression in pistil, which may both be related to floral development.

Figure 3. Analysis of flower development and anthocyanin biosynthesis related genes *Brasenia schreberi*.
(a) Photograph of *B. schreberi* flower.
(b) Expression patterns of type II MADS-box genes from various organs (from left to right: root, stem, mature leaf, juvenile leaf, pistil, stamen, petal and sepal) of *B. schreberi*. Expression values were scaled by $\log_2$(FPKM+1), in which FPKM is fragments per kilobase of exon per million mapped reads.
(c) Relative contents of anthocyanins at six organizations.
(d) Gene expression levels ($\log_{10}$(FPKM+1)) at different tissues (stamen, pistil, sepal, petal, mature leaf, and juvenile leaf) are represented by color grading. Gene IDs are shown on the right side of each heatmap. Genes identified in tandem clusters are marked in turquois color. The leaf anthocyanins and flavonoid glycosides are marked and framed. *4CL*, 4-coumarate CoA ligase 4; *ANR*, anthocyanidin reductase; *ANS*, anthocyanidin synthase; *C4H*, cinnamate-4-hydroxylase; *CHI*, chalcone isomerase; *CHS*, chalcone synthase; *DFR*, dihydroflavonol 4-reductase; *F3′5′H*, flavonoid 3′5′-hydroxylase; *F3′H*, flavonoid 3′-hydroxylase; *F3H*, flavanone 3-hydroxylase; *FLS*, flavonol synthase; *PAL*, phenylalanine ammonia-lyase; *UFGT*, flavonoid-3-O-glucosyltransferase.

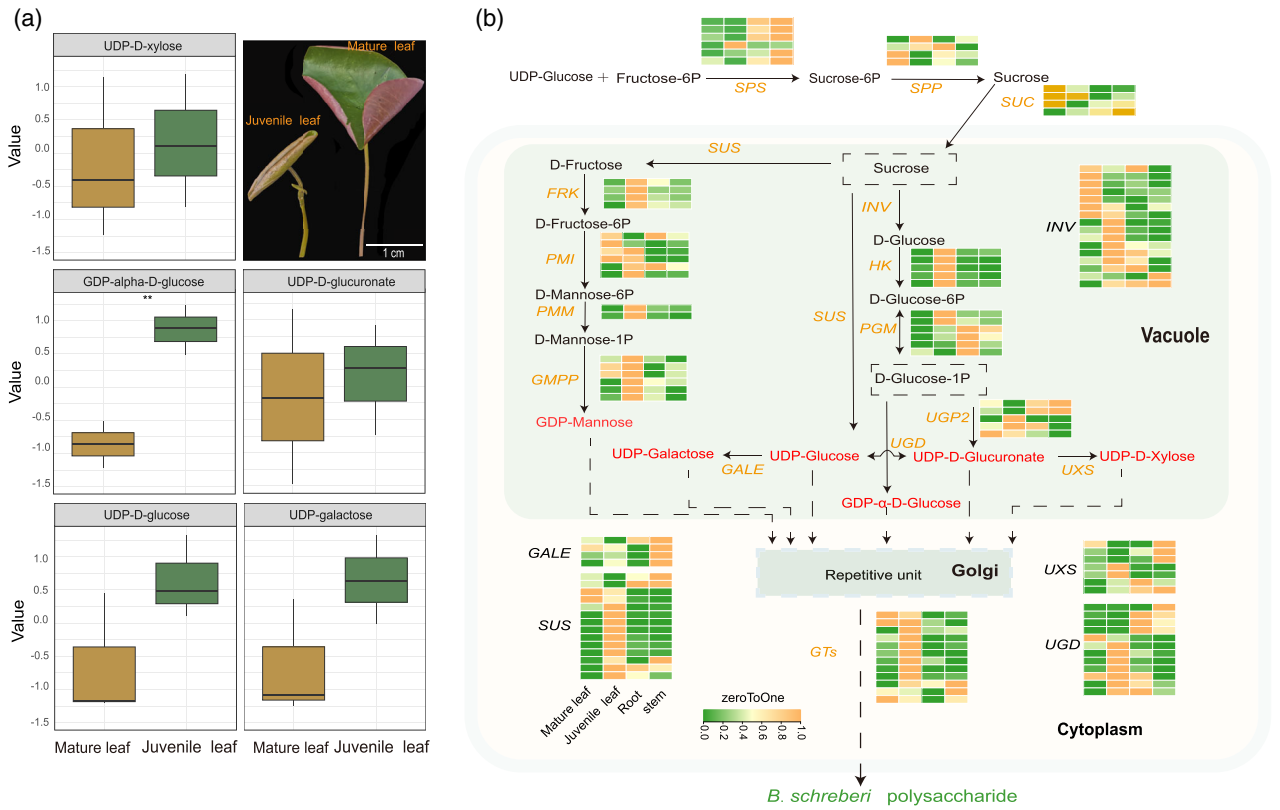

Figure 4. *Brasenia schreberi* polysaccharide biosynthesis pathway and sugar transporter protein transcription factor identification.
(a) Biomass of five polysaccharide monomers in mature leaf and juvenile leaf.
(b) Gene expression analysis of genes involved in polysaccharide biosynthesis in *B. schreberi*. Different color blocks represent the expression levels of the encoded genes in different tissues; the squares from left to right correspond to mature leaf, juvenile leaf, root and stem. Detailed gene names can be found in Table S14.

In addition to the MADS-box genes, we further investigated the pigments' accumulation in the different flower organs. Pigments, responsible for the majority of coloration in angiosperms, are flavonoid derivatives, predominantly sourced from three precursors in the anthocyanin pathway: pelargonidin (red flowers), cyanidin (ranging from red to purple to blue), and delphinidin (blue or purple flowers) (Wessinger & Rausher, 2012). The flowers of *B. schreberi* display a uniform purplish-red color (Figure 3a). By analyzing the transcriptome and metabolites of *B. schreberi*, we constructed a metabolic pathway for its flavonoids and anthocyanins (Figure 3d). We discovered that only two precursors appeared in the anthocyanin pathway of *B. schreberi*, including delphinidin and cyanidin. The content of cyanidin and delphinidin demonstrates the basic structure of the flower, with no significant differences in anthocyanin content among petals, sepals, stamens, and pistils, corroborating the observation of a single purplish-red color in the flowers (Figure 3a,c). We also delved into the pigments present in the leaves of *B. schreberi*. Notably, the upper side of mature leaves that float on the water surface exhibits a greenish-blue hue, whereas the underside submerged in water displays a purplish-red color, with the submerged juvenile leaves appearing tender green. Our investigation revealed that the anthocyanin content is higher in mature leaves compared to juvenile leaves (Figure 3c). Moreover, we observed that the expression levels of enzymes involved in the anthocyanin synthesis pathway are higher in juvenile leaves than in mature leaves, particularly for the four key genes *C4H*, *DFR*, *ANR*, and *ANS* (Figure 3d). This differential expression pattern may originate from more active anthocyanin biosynthesis during juvenile leaf growth, followed by progressive accumulation during development, ultimately leading to higher anthocyanin content in mature leaves than in juvenile leaves. This phenomenon demonstrates strong concordance with their distinctive phenotypic characteristics (Figure 4a).

Mucilage synthesis in *B. schreberi* juvenile leaves

Mucilage, a viscous polysaccharide-rich secretion produced by specialized epidermal cells, forms a protective coating on plant surfaces that serves as an effective barrier against microbial colonization in aquatic habitats

(Galloway et al., 2020). The submerged juvenile leaves of *B. schreberi* are wrapped in a thick layer of transparent gelatinous mucilage, which is secreted by mucilage hairs, rich in polysaccharides (Ai et al., 2024; Xie et al., 2018). However, this mucilage disappears when juvenile leaves develop into mature leaves, which are on the water surface. Comparative transcriptomic analysis between these two distinct leaf developmental stages revealed 3569 differentially expressed genes (DEGs), comprising 1556 upregulated and 2014 downregulated transcripts in juvenile leaves relative to mature leaves (Figure S10). Functional enrichment analysis demonstrated that mature leaf-upregulated genes were predominantly associated with photosynthetic apparatus organization (GO:0009521), chloroplast membrane biogenesis (GO:0009706), and carbohydrate metabolism (GO:0046364) (Table S13). Conversely, juvenile leaf-enriched transcripts were significantly overrepresented in biological processes related to polysaccharide biosynthesis, including transmembrane water transport (GO:0005372), glucosyl transfer reactions (GO:0046527), and sucrose metabolism (GO:0016157) (Table S13), consistent with the observed mucilage production in developing leaves.

Our comprehensive analysis of the polysaccharide biosynthetic pathway in *B. schreberi* revealed an integrated metabolic network comprising sucrose synthesis, transport, and subsequent conversion to polysaccharide monomers, involving 112 functionally annotated genes (Figure 4b) (Cui et al., 2025). The key sucrose synthesis enzyme sucrose-6-phosphate phosphohydrolase (*SPP*) (Mehdi et al., 2024) exhibited consistently high expression levels across both mature and juvenile leaves, demonstrating their shared capacity for sucrose production. This sucrose pool is subsequently mobilized by four highly expressed sucrose transporter genes (*SUC*s) showing ubiquitous expression patterns throughout all examined tissues. Enzymatic conversion of sucrose yielded diverse monosaccharide intermediates including GDP-mannose, UDP-galactose, UDP-glucose, UDP-D-glucuronate, UDP-D-xylose, and GDP-glucose, with metabolic profiling revealing GDP-glucose as the only differentially accumulated metabolite between leaf developmental stages. This metabolic signature correlated strongly with the juvenile leaf-enriched expression profile of GDP-glucose biosynthetic enzymes (*INV*, *HK*, and *PGM*), particularly the five *HK* gene copies showing developmentally restricted expression. Additional monosaccharide biosynthesis enzymes displayed either juvenile leaf-preferential or constitutive expression patterns (Figure 4a). The terminal glycosylation steps, mediated by glycosyltransferases (*GT*s) (Lairson et al., 2008; Zhang, Chen, et al., 2020; Zhang, Zhang, et al., 2020), were predominantly associated with elevated *GT* expression in juvenile leaves, consistent with active polysaccharide biosynthesis during early leaf development.

CONCLUSIONS

Our study presents a high-quality T2T genome assembly of *B. schreberi*, achieved through integrated PacBio HiFi, ONT long-read, and Hi-C sequencing technologies. We identified two WGD events and one WGT in this early-diverging aquatic angiosperm, with phylogenetic analysis confirming the most ancient WGD as a shared genomic feature within Nymphaeaceae. Our findings demonstrate that these large-scale genomic alterations, combined with DSDs, facilitated the expansion of gene families associated with environmental adaptation. Through transcriptomic and metabolomic analyses, we uncovered key metabolic pathways involved in flavonoid and anthocyanin biosynthesis, while revealing differential gene expression patterns correlated with mucilage distribution, which represents a crucial adaptive trait in this aquatic species. Our results further show that mature leaves conduct photosynthesis and likely transport nutrients to juvenile leaves to support polysaccharide biosynthesis and mucilage production. Additionally, we established that uniform expression of anthocyanin biosynthesis genes across floral organs results in the characteristic purple pigmentation of *B. schreberi* flowers. Collectively, our work provides fundamental insights into the genomic basis of chromosomal evolution and ecological adaptation in aquatic angiosperms, significantly advancing understanding of genome dynamics and adaptive strategies within Nymphaeales.

METHODS

Plant materials

The plant material for this study was obtained from water shield plants in May 2023. The sampling site was located in Maqian Village, Lichuan City, Enshi Tujia and Miao Autonomous Prefecture, Hubei Province, China (latitude 30°16′49″ N, longitude 108°50′03″ E). Eight different tissue types were harvested: sepal, petal, stamen, pistil, root, stem, leaf, and juvenile leaf. Fresh leaf tissue was used for DNA extraction to facilitate subsequent genome sequencing. At the same time, samples from all eight tissue types were collected and preserved for RNA extraction, in which calyx, petal, stamen, pistillate, leaf, and juvenile leaf were analyzed for comprehensive metabolomics, with the former serving as the basis for transcriptomic analysis. This strategic sampling method allows for a multifaceted examination of the molecular and biochemical characteristics of plants at different organs and stages of development.

Library construction and sequencing

For PacBio HiFi sequencing, SMRTbell target-size libraries were constructed for sequencing following PacBio's standard protocol (Pacific Biosciences, CA, USA), utilizing 15 kb preparation solutions. In brief, 15 μg of genomic DNA per sample was used for library preparation. The DNA was sheared using g-TUBEs (Covaris, MA, USA) to achieve the desired fragment size for the library. Subsequently, single-strand overhangs were removed, and DNA fragments were subjected to damage repair, end repair, and A-tailing. The fragments were then ligated to the hairpin adaptor for

PacBio sequencing. The library underwent treatment with the SMRTbell Enzyme Cleanup Kit, followed by purification with AMPure PB Beads. Target fragments were selected using the BluePippin system (Sage Science, MA, USA). The SMRTbell library was further purified using AMPure PB beads, and the size of the library fragments was assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA). Sequencing was conducted on a PacBio Sequel II instrument using Sequencing Primer V2 and Sequel II Binding Kit 2.1 at Haorui Genomics, Xian, China.

For Oxford Nanopore sequencing, fresh leaves were collected from *B. schreberi*. High-quality genomic DNA was isolated from the fresh leaves using the CTAB method, and the DNA quality and concentration were tested by 0.75% agarose gel electrophoresis, NanoDrop One spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and Qubit 3.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). The libraries were prepared using the SQKLSK109 ligation kit and the standard protocol. The purified library was loaded onto primed R9.4 Spot-On Flow Cells and sequenced using a PromethION sequencer (Oxford Nanopore Technologies, Oxford, UK) with 48-h runs at Wuhan Benagen Technology Co. Ltd., Wuhan, China. Base calling analysis of raw data was performed using the Oxford Nanopore GUPPY software (v0.3.0).

In addition, Hi-C (high-throughput chromosome conformation capture) sequencing was performed. Fresh leaves were fixed in 1% formaldehyde solution for chromatin cross-linking, followed by digestion with the restriction enzyme Dpn II. The resulting DNA was used to construct a library, which was then sequenced using the Illumina HiSeq 4000 platform.

Genome size and heterozygosity estimation

Initially, we utilized Jellyfish v.2.10 (Marçais & Kingsford, 2011) to calculate the depth distribution of K-mers, with a K-mer size of 19. Then, GenomeScope v.1.0.0 (Vurture et al., 2017) was employed to model the K-mer spectrum, which facilitated the prediction of genomic characteristics such as genome size and heterozygosity.

Genome assembly and pseudochromosome construction

Upon acquiring high-accuracy continuous long-read (CCS) data, Oxford Nanopore (ONT) ultra-long reads and Hi-C data, we conducted a contig-level genome assembly utilizing hifiasm (v 0.19.5) (Cheng et al., 2021). Following this, the genomic completeness was evaluated with BUSCO (v5.4.3) (Simão et al., 2015) and Merqury (v 1.0) (Rhie et al., 2020). For the BUSCO assessment, we employed the lineage-specific gene database, embryophyta_odb10, to ensure precise evaluation of genome integrity.

Hi-C data were utilized for chromosome-scale scaffolding. Initially, we employed HiC-Pro (v.3.1.0) (Servant et al., 2015) to filter the Hi-C data digested by Mbol, resulting in clean Hi-C data. Subsequently, these clean Hi-C data were aligned to the assembled genome using Juicer (v.1.6) (Durand et al., 2016). Chromosome-level assembly was conducted using 3D-DNA (v.180922) (Dudchenko et al., 2017), which involved the clustering, sorting, and orienting of genomic proximity information derived from Hi-C read pairs. The Hi-C cross-linking signals were then visualizedusing Juicebox (v.1.11.08), linking adjacent anchored assemblies to form 12 super assemblies corresponding to 36 pseudo-chromosomes. Lastly, gaps within the chromosome-level genome were sequentially filled using TGS-GapCloser (v.1.2.1) (Xu et al., 2020) and quarTeT (v.1.1.4) (Lin et al., 2023).

Identification of telomeres and centromeres

The TeloExplore module of quarTeT software v1.1.4 (Lin et al., 2023) was used to identify telomeres in the *B. schreberi* genome with the parameter '-c plant'. Telomeric repeats were identified using TIDK v0.2.63 (Brown, 2025) with the following parameters: 'explore --minimum 5 --maximum 15 -t 8 --log'. Given that telomeric regions are repetitive, the unit 'AAAGGATAAGAC', which we use in this article, was employed. Telomere locations were further refined using the TIDK search function with the option -s 'AAAGGATAAGAC'.

For centromere identification, the CentroMiner module of quarTeT software was employed with default parameters. To visualize the position of repeat units in the *B. schreberi* genome, a bed file was created using the awk command.

Repeat element identification and gene prediction

RepeatMasker (Tarailo-Graovac & Chen, 2009) and RepeatProteinMasker (Tarailo-Graovac & Chen, 2009) were used to identify repetitive elements based on homology alignments between *B. schreberi* genome sequences and Repbase (v.16.10). We then applied the *de novo* approach to improve the sensitivity of our repeat identification. Briefly, RepeatModeler (Bao et al., 2015) and LTR_Finder (Xu & Wang, 2007) (v1.06) were selected to construct the repeat library, and then RepeatMasker (Tarailo-Graovac & Chen, 2009) was employed to generate the *de novo* predictions.

We employed three strategies to predict and annotate genes in the *B. schreberi* genome, including *de novo*, homology-based, and transcriptome-based methods. For *de novo* prediction, we used GlimmerHMM (v.3.0.4) (Majoros et al., 2004), Augustus (v.3.0.2) (Stanke & Morgenstern, 2005), and Genscan (v.1.0) to predict genes. In the case of homology-based prediction, we aligned protein sequences from *E. ferox, N. colorata, Nymphaea thermarum* (Povilus et al., 2020), *B. schreberi* (Lu et al., 2023), *V. vinifera* (Zheng et al., 2021) to the *B. schreberi* genome using tBlastn (Altschul, 1997) (*E*-value $<1e^{-5}$), and then used Genewise (v.2.4.1) to annotate the genes based on these alignments. For transcriptome-based prediction, RNA-seq data was aligned to the genome using Hisat2 (Kim et al., 2019) (v.2.2.1) and assembled into gene models using Cufflinks (Trapnell et al., 2012) (v.2.2.1). Finally, with the assistance of Evidence Modeler (Haas et al., 2008) (EVM, v.1.11), all genes predicted by these various methods were integrated into a non-redundant and complete gene set. Functional annotation of protein-coding genes was conducted by running eggNOG-mapper (v.2.1.6) (Cantalapiedra et al., 2021) and Blast (Altschul, 1997) (v.2.2.3.1) (*E*-value $\leq 1e^{-5}$) against the NCBI non-redundant (Nr) protein database, Swiss-Prot (Bairoch, 2000), Pfam (Mistry et al., 2021), and KEGG (Kanehisa, 2000).

Genome evolution and expansion/contraction of gene families

To investigate the evolutionary trajectory of the *B. schreberi* genome, a total of 17 other species were selected for phylogenetic analysis: *G. biloba* (Liu et al., 2021), *V. vinifera* (Velasco et al., 2007), *A. trichopoda* (Amborella Genome Project et al., 2013), *A. gramineus* (Guo et al., 2023), *N. nucifera* (Li et al., 2021), *E. ferox* (Ma et al., 2021), *V. cruziana* (Wen et al., 2025), *N. colorata* (Zhang, Chen, et al., 2020) and *B. schreberi*. An all-versus-all BLASTP (Altschul, 1997) (v2.2.26) (*E*-value cut-off: 1×10^{-5}) was first employed to generate similarity information for all genes. Then, we identified high-quality single-copy genes by applying SonicParanoid v1.0 (Cosentino & Iwasaki, 2019) and constructed a phylogenetic tree with this gene set using IQ-TREE

(Nguyen et al., 2015) (v1.6.12). We further estimated the times of divergence between species with MCMCtree in the PAML package (v4). The divergence times of *A. thaliana* and *V. vinifera* (109–124 Mya) and those of *G. biloba* and *A. trichopoda* (179–205 Mya) obtained from TimeTree (http://www.timetree.org) were used as calibration points. Gene family expansion and contraction were further estimated by CAFÉ (De Bie et al., 2006) (v4.2) using the gene cluster information and the estimated time tree. The parameter λ was estimated along each branch with the random model, and then all the gene families were classified into three types: expanded, contracted, or unchanged. The final phylogenetic tree was drawn using the online tool Chiplot (https://www.chiplot.online/) and subsequently refined in Adobe Illustrator (v 26.0.1).

WGD and subgenome phylogeny

Utilizing the Whole-Genome Duplication Integration (WGDI) software (Sun et al., 2022, 2024), we selected species from Nymphaeaceae as the reference genome to elucidate the WGD events in *B. schreberi*. For subgenome delineation, the Angiosperms Ancestral Karyotype (AAk) was selected as the outgroup for all species. The specific methods included: using '-d' and '-icl' parameters to identify collinear genes and homologous blocks between two species; employing '-ks', 'kp', and 'pf' parameters to calculate synonymous substitution rates (*Ks*), plot *Ks* distributions, and fit *Ks* peaks; using '-km' parameter to map ancestral karyotypes onto the studied species followed by subgenome partitioning; and finally extracting genes from each subgenome set using '-pc', '-a', and '-at' parameters to construct phylogenetic trees. Then, we integrated these gene trees using the multi-locus coalescent tree-building software Astral (v5.7.8).

Transcriptome sequencing and analysis of DEGs

The samples came from eight different tissues of the same plant, including petals, sepal, stamen, pistil, mature leaf, juvenile leaf, root, and stems, with three biological copies of each sample. RNA extraction and cDNA library construction were performed. Filtered clean reads were mapped to the *B. schreberi*. genome using Hisat2 (Kim et al., 2019) (v.2.0.4). Using StringTie (Pertea et al., 2016) (v.2.2.1) performed transcript assembly to obtain gene expression levels and reads counts for each gene. We used DESeq2 software package to analyze DEGs, and defined the genes with log2-fold change (FC) >1, *P* value ≤0.001, and false discovery rate (FDR) <0.05 as significant genes.

Flower development genes analysis

MADS-box genes were identified using the HMMER (Finn et al., 2011) and iTAK (Zheng et al., 2016) software, and the parameters '-cut_tc' and Pfam profiles (PF00319) were used for HMM searching. For the other FDRGs, we performed the protein sequences similarity search by BLASTP with an *E*-value cut-off of 10^{-5} using the known flowering genes in *A. thaliana* as a reference. InterProScan was applied to further check the integrity of candidate gene domains. Multiple sequence alignment and ML tree inference were performed to group them into subfamilies, and the gene set that clustered with the reference sequence was used for the next analysis.

Polysaccharide synthesis in *B. schreberi*

Genes associated with polysaccharide synthesis were initially identified through the KEGG database. Subsequently, polysaccharide-related genes in *B. schreberi* were screened using HMMER (Finn et al., 2011) and BLASTP. In the HMMER analysis, the parameters '-cut_tc' and Pfam profiles (e.g., PF00319) from the Pfam database were employed for Hidden Markov Model (HMM) searching. For other candidate genes, protein sequence similarity searches were conducted using BLASTP with an *E*-value cut-off of 10^{-5}, using known carbohydrate synthesis-related genes in *A. thaliana* as a reference. The results above were combined to determine the genes related to the polysaccharide pathway of *B. schreberi*.

ACKNOWLEDGEMENTS

The project was supported by the National Natural Science Foundation of China (No. 32201588 to Huiying Shang and No. 32170219 to Yongzhi Yang). We would also like to thank Jin Zhang, Liqiang Hou, and Shuai Yuan for their help with the methodology.

AUTHOR CONTRIBUTIONS

YY and HS conceived and supervised the project. CL and HS performed sample collection and conducted the majority of data analyses. HS and CL wrote the original manuscript draft with critical input from YY. All authors contributed to manuscript revisions and approved the final version.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data supporting this work are available within the paper and its supplemental information. Raw reads for *B. schreberi* have been deposited at CNCB under the BioProject accession numbers of PRJCA040856. Genome assemblies, annotations and metabolism files are available at Figshare (https://doi.org/10.6084/m9.figshare.29178863).

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Genome size estimation by GenomeScope.

Figure S2. Interaction frequency distribution of Hi-C links among chromosomes.

Figure S3. The collinear relationship of V4_T2T, V1, and V3 assemblies.

Figure S4. Interaction frequency distribution of Hi-C links among chromosomes 1 and 17 in versions V1 and V4_T2T.

Figure S5. Completeness assessment and whole proteome assessment.

Figure S6. K-mer spectra analysis of the *Brasenia* genome assembly.

Figure S7. Collinear dot map of the *B. schreberi* genome.

Figure S8. Intra- and inter-specific *Ks* dot plots of *N. colorata, B. schreberi*, and *E. ferox*.

Figure S9. (a) Cluster dendrogramm of transcriptomes. (b) OrthogonalPartialLeast Squares-DiscriminantAnalysis of metabolomes.

Figure S10. Volcano map of differentially expressed genes in mature and juvenile leaves, with juvenile leaves as the control.

Figure S11. Analysis of Type I of MADS-box genes in five species.

Figure S12. Analysis of Type II MADS-box genes in five species.

Figure S13. Analysis of ERF genes fimaly in five species.

Table S1. Data statistics of whole-genome sequencing.

Table S2. Length and gap number comparison between V4_T2T, V3, and V1.

Table S3. Summary of *Brasenia scherberi* genome (V4_T2T) telomeres.

Table S4. Summary of *Brasenia scherberi* genome (V4_T2T) assembly and annotation.

Table S5. Functional prediction of encoded proteins (V4_T2T).

Table S6. Function enrichment of *Brasenia schreberi* expansion genes.

Table S7. Function enrichment of *Brasenia schreberi* contraction genes.

Table S8. Statistics on the number of genes that expand and contract at each node in a phylogenetic tree.

Table S9. The replication types of expanded gene families in the ancestral lineage of Nymphaeales.

Table S10. The list of MADS-Box genes in *Brasenia schreberi*, *Nymphaea colorata*, *Arabiopsis thaliana*, *Oryza sativa*, and *Amborella trichopoda*.

Table S11. Anthocyanin content of *Brasenia schreberi* in six tissues.

Table S12. The list of anthocyanin synthesis pathway genes in *Brasenia schreberi*.

Table S13. Functional enrichment of differentially expressed genes in mature and juvenile leaves.

Table S14. The list of i polysaccharide biosynthesis pathway genes in *Brasenia schreberi*.

Table S15. Function enrichment of Nymphaeales expansion genes.

Table S16. Function enrichment of *Nymphaea colorata* expansion genes.

Table S17. Function enrichment of *Euryale ferox* expansion genes.

REFERENCES

Ai, T., **Liu, H.**, **Wan, J.**, **Lu, B.**, **Yu, X.**, **Liu, J.** ***et al.*** (2024) Proteomics analysis reveals the underlying factors of mucilage disappearance in *Brasenia schreberi* and its influence on nutrient accumulation. *Food*, **13**, 518.

Altschul, S. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389–3402.

Amborella Genome Project, **Albert, V.A.**, **Barbazuk, W.B.**, **dePamphilis, C.W.**, **Der, J.P.**, **Leebens-Mack, J.** ***et al.*** (2013) The *Amborella* genome and the evolution of flowering plants. *Science*, **342**, 1241089.

Anderson, J.T., **Willis, J.H.** & **Mitchell-Olds, T.** (2011) Evolutionary genetics of plant adaptation. *Trends in Genetics*, **27**, 258–266.

Arora, R., **Agarwal, P.**, **Ray, S.**, **Singh, A.K.**, **Singh, V.P.**, **Tyagi, A.K.** ***et al.*** (2007) MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genomics*, **8**, 242.

Bairoch, A. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Research*, **28**, 45–48.

Bao, W., **Kojima, K.K.** & **Kohany, O.** (2015) Repbase update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, **6**, 11.

Bowles, A.M.C., **Williamson, C.J.**, **Williams, T.A.**, **Lenton, T.M.** & **Donoghue, P.C.J.** (2023) The origin and early evolution of plants. *Trends in Plant Science*, **28**, 312–329.

Brown, M.R., **Manuel Gonzalez de La Rosa, P.** & **Blaxter, M.** (2025) tidk: a toolkit to rapidly identify telomeric repeats from genomic datasets. *Bioinformatics*, **41**, btaf049.

Cantalapiedra, C.P., **Hernández-Plaza, A.**, **Letunic, I.**, **Bork, P.** & **Huerta-Cepas, J.** (2021) eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular Biology and Evolution*, **38**, 5825–5829.

Chanderbali, A.S., **Albert, V.A.**, **Ashworth, V.E.T.M.**, **Clegg, M.T.**, **Litz, R.E.**, **Soltis, D.E.** ***et al.*** (2008) *Persea americana* (avocado): bringing ancient flowers to fruit in the genomics era. *BioEssays*, **30**, 386–396.

Cheng, H., **Concepcion, G.T.**, **Feng, X.**, **Zhang, H.** & **Li, H.** (2021) Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods*, **18**, 170–175.

Cosentino, S. & **Iwasaki, W.** (2019) SonicParanoid: fast, accurate and easy orthology inference. *Bioinformatics*, **35**, 149–151.

Cui, J., **Wang, R.**, **Gu, R.**, **Chen, M.**, **Wang, Z.**, **Li, L.** ***et al.*** (2025) Telomere-to-telomere *Phragmites australis* reference genome assembly with a B chromosome provides insights into its evolution and polysaccharide biosynthesis. *Communications Biology*, **8**, 1–20.

De Bie, T., **Cristianini, N.**, **Demuth, J.P.** & **Hahn, M.W.** (2006) CAFE: a computational tool for the study of gene family evolution. *Bioinformatics*, **22**, 1269–1271.

Dudchenko, O., **Batra, S.S.**, **Omer, A.D.**, **Nyquist, S.K.**, **Hoeger, M.**, **Durand, N.C.** ***et al.*** (2017) De novo assembly of the *Aedes aegypti* genome using hi-C yields chromosome-length scaffolds. *Science*, **356**, 92–95.

Durand, N.C., **Shamim, M.S.**, **Machol, I.**, **Rao, S.S.P.**, **Huntley, M.H.**, **Lander, E.S.** ***et al.*** (2016) Juicer provides a one-click system for analyzing loop-resolution hi-C experiments. *Cell Systems*, **3**, 95–98.

Fajkus, J., **Sýkorová, E.** & **Leitch, A.R.** (2005) Telomeres in evolution and evolution of telomeres. *Chromosome Research*, **13**, 469–479.

Finn, R.D., **Clements, J.** & **Eddy, S.R.** (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*, **39**, W29–W37.

Galloway, A.F., **Knox, P.** & **Krause, K.** (2020) Sticky mucilages and exudates of plants: putative microenvironmental design elements with biotechnological value. *The New Phytologist*, **225**, 1461–1469.

Gigord, L.D., **Macnair, M.R.** & **Smithson, A.** (2001) Negative frequency-dependent selection maintains a dramatic flower color polymorphism in the rewardless orchid *Dactylorhiza sambucina* (L.) Soo. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 6253–6255.

Guo, X., **Wang, F.**, **Fang, D.**, **Lin, Q.**, **Sahu, S.K.**, **Luo, L.** ***et al.*** (2023) The genome of *Acorus* deciphers insights into early monocot evolution. *Nature Communications*, **14**, 3662.

Haas, B.J., **Salzberg, S.L.**, **Zhu, W.**, **Pertea, M.**, **Allen, J.E.**, **Orvis, J.** ***et al.*** (2008) Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biology*, **9**, R7.

Hongyi, C., **Mei, S.U.N.**, **Zhenya, L.I.U.** & **Hangmei, Y.** (2024) Response of leaf economic traits of natural *Brasenia schreberi* to water environment in high altitude area. *Chinese Journal of Ecology*, **43**, 1763.

Hopkins, R. (2013) Reinforcement in plants. *The New Phytologist*, **197**, 1095–1103.

Kanehisa, M. (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, **28**, 27–30.

Kim, D., **Paggi, J.M.**, **Park, C.**, **Bennett, C.** & **Salzberg, S.L.** (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, **37**, 907–915.

Lairson, L.L., **Henrissat, B.**, **Davies, G.J.** & **Withers, S.G.** (2008) Glycosyltransferases: structures, functions, and mechanisms. *Annual Review of Biochemistry*, **77**, 521–555.

Li, G., **Zhao, X.**, **Yang, J.**, **Hu, S.**, **Ponnu, J.**, **Kimura, S.** ***et al.*** (2024) Water wisteria genome reveals environmental adaptation and heterophylly regulation in amphibious plants. *Plant, Cell & Environment*, **47**, 4720–4740.

Li, H., **Yang, X.**, **Zhang, Y.**, **Gao, Z.**, **Liang, Y.**, **Chen, J.** ***et al.*** (2021) *Nelumbo* genome database, an integrative resource for gene expression and variants of *Nelumbo nucifera*. *Scientific Data*, **8**, 38.

Lin, Y., **Ye, C.**, **Li, X.**, **Chen, Q.**, **Wu, Y.**, **Zhang, F.** ***et al.*** (2023) quarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. *Horticulture Research*, **10**, uhad127.

Liu, H., **Wang, X.**, **Wang, G.**, **Cui, P.**, **Wu, S.**, **Ai, C.** ***et al.*** (2021) The nearly complete genome of *Ginkgo biloba* illuminates gymnosperm evolution. *Nature Plants*, **7**, 748–756.

Lu, B., **Shi, T.** & **Chen, J.** (2023) Chromosome-level genome assembly of watershield (*Brasenia schreberi*). *Scientific Data*, **10**, 467.

Ma, J., **Sun, P.**, **Wang, D.**, **Wang, Z.**, **Yang, J.**, **Li, Y.** ***et al.*** (2021) The *Chloranthus sessilifolius* genome provides insight into early diversification of angiosperms. *Nature Communications*, **12**, 6929.

Majoros, W.H., **Pertea, M.** & **Salzberg, S.L.** (2004) TigrScan and GlimmerHMM: two open source *ab initio* eukaryotic gene-finders. *Bioinformatics*, **20**, 2878–2879.

Marçais, G. & **Kingsford, C.** (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, **27**, 764–770.

Mehdi, F., **Galani, S.**, **Wickramasinghe, K.P.**, **Zhao, P.**, **Lu, X.**, **Lin, X.** ***et al.*** (2024) Current perspectives on the regulatory mechanisms of sucrose accumulation in sugarcane. *Heliyon*, **10**, e27277.

Mistry, J., **Chuguransky, S.**, **Williams, L.**, **Qureshi, M.**, **Salazar, G.A.**, **Sonnhammer, E.L.L.** ***et al.*** (2021) Pfam: the protein families database in 2021. *Nucleic Acids Research*, **49**, D412–D419.

Nguyen, L.-T., **Schmidt, H.A.**, **Von Haeseler, A.** & **Minh, B.Q.** (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, **32**, 268–274.

Pertea, M., **Kim, D.**, **Pertea, G.M.**, **Leek, J.T.** & **Salzberg, S.L.** (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, **11**, 1650–1667.

Povilus, R.A., **DaCosta, J.M.**, **Grassa, C.**, **Satyaki, P.R.V.**, **Moeglein, M.**, **Jaenisch, J.** ***et al.*** (2020) Water lily (*Nymphaea thermarum*) genome reveals variable genomic signatures of ancient vascular cambium losses. *Proceedings of the National Academy of Sciences*, **117**, 8649–8656.

Qiu, Y., **Li, Z.** & **Köhler, C.** (2024) Ancestral duplication of MADS-box genes in land plants empowered the functional divergence between sporophytes and gametophytes. *The New Phytologist*, **244**, 358–363.

Rhie, A., **Walenz, B.P.**, **Koren, S.** & **Phillippy, A.M.** (2020) Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology*, **21**, 245.

Sasidharan, R. & **Mustroph, A.** (2011) Plant oxygen sensing is mediated by the N-end rule pathway: a milestone in plant anaerobiosis. *Plant Cell*, **23**, 4173–4183.

Servant, N., **Varoquaux, N.**, **Lajoie, B.R.**, **Viara, E.**, **Chen, C.-J.**, **Vert, J.-P.** ***et al.*** (2015) HiC-pro: an optimized and flexible pipeline for hi-C data processing. *Genome Biology*, **16**, 259.

Simão, F.A., **Waterhouse, R.M.**, **Ioannidis, P.**, **Kriventseva, E.V.** & **Zdobnov, E.M.** (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**, 3210–3212.

Stanke, M. & **Morgenstern, B.** (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research*, **33**, W465–W467.

Sun, P., **Jiao, B.**, **Yang, Y.**, **Shan, L.**, **Li, T.**, **Li, X.** ***et al.*** (2022) WGDI: a user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. *Molecular Plant*, **15**, 1841–1851.

Sun, P., **Lu, Z.**, **Wang, Z.**, **Wang, S.**, **Zhao, K.**, **Mei, D.** ***et al.*** (2024) Subgenome-aware analyses reveal the genomic consequences of ancient allopolyploid hybridizations throughout the cotton family. *Proceedings of the National Academy of Sciences*, **121**, e2313921121.

Tarailo-Graovac, M. & **Chen, N.** (2009) Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*, **25**, 4.10.1–4.10.14.

Titus, J.E. & **Urban, R.A.** (2009) Aquatic plants: a general introduction. In: Likens, G.E. (Ed.) *Encyclopedia of inland waters*. Oxford: Academic Press, pp. 43–51. Available from: https://www.sciencedirect.com/science/article/pii/B9780123706263002143 (Accessed February 14, 2025)

Trapnell, C., **Roberts, A.**, **Goff, L.**, **Pertea, G.**, **Kim, D.**, **Kelley, D.R.** ***et al.*** (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nature Protocols*, **7**, 562–578.

Velasco, R., **Zharkikh, A.**, **Troggio, M.**, **Cartwright, D.A.**, **Cestaro, A.**, **Pruss, D.** ***et al.*** (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS One*, **2**, e1326.

Vurture, G.W., **Sedlazeck, F.J.**, **Nattestad, M.**, **Underwood, C.J.**, **Fang, H.**, **Gurtowski, J.** ***et al.*** (2017) GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics*, **33**, 2202–2204.

Wang, Y., **Zou, Y.**, **Fang, Q.**, **Feng, R.**, **Zhang, J.**, **Zhou, W.** ***et al.*** (2023) Polysaccharides from *Brasenia schreberi* with great antioxidant ability and the potential application in yogurt. *Molecules*, **29**, 150.

Wei, X., **Xu, H.**, **Rong, W.**, **Ye, X.** & **Zhang, Z.** (2019) Constitutive expression of a stabilized transcription factor group VII ethylene response factor enhances waterlogging tolerance in wheat without penalizing grain yield. *Plant, Cell & Environment*, **42**, 1471–1485.

Wen, X., **Liang, Y.**, **Shan, H.**, **Chang, X.**, **Song, X.**, **Shen, S.** ***et al.*** (2025) The genome of giant waterlilies provides insights into the origin of angiosperms, leaf gigantism, and stamen function innovation. *Plant Communications*, **6**, 101342.

Wessinger, C.A. & **Rausher, M.D.** (2012) Lessons from flower colour evolution on targets of selection. *Journal of Experimental Botany*, **63**, 5741–5749.

Wu, P., **Zhang, L.**, **Zhang, K.**, **Yin, Y.**, **Liu, A.**, **Zhu, Y.** ***et al.*** (2022) The adaptive evolution of *Euryale ferox* to the aquatic environment through paleo-hexaploidization. *The Plant Journal*, **110**, 627–645.

Xiao, H., **Cai, X.**, **Fan, Y.** & **Luo, A.** (2016) Antioxidant activity of water-soluble polysaccharides from *Brasenia schreberi*. *Pharmacognosy Magazine*, **12**, 193–197.

Xie, C., **Li, J.**, **Pan, F.**, **Fu, J.**, **Zhou, W.**, **Lu, S.** ***et al.*** (2018) Environmental factors influencing mucilage accumulation of the endangered *Brasenia schreberi* in China. *Scientific Reports*, **8**, 17955.

Xu, M., **Guo, L.**, **Gu, S.**, **Wang, O.**, **Zhang, R.**, **Peters, B.A.** ***et al.*** (2020) TGS-GapCloser: a fast and accurate gap closer for large genomes with low coverage of error-prone long reads. *GigaScience*, **9**, giaa094.

Xu, Z. & **Wang, H.** (2007) LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research*, **35**, W265–W268.

Yang, Y., **Sun, P.**, **Lv, L.**, **Wang, D.**, **Ru, D.**, **Li, Y.** ***et al.*** (2020) Prickly waterlily and rigid hornwort genomes shed light on early angiosperm evolution. *Nature Plants*, **6**, 215–222.

Yu, F., **Liang, K.**, **Fang, T.**, **Zhao, H.**, **Han, X.**, **Cai, M.** ***et al.*** (2019) A group VII ethylene response factor gene, ZmEREB180, coordinates waterlogging tolerance in maize seedlings. *Plant Biotechnology Journal*, **17**, 2286–2298.

Zhang, L., **Chen, F.**, **Zhang, X.**, **Li, Z.**, **Zhao, Y.**, **Lohaus, R.** ***et al.*** (2020) The water lily genome and the early evolution of flowering plants. *Nature*, **577**, 79–84.

Zhang, P., **Zhang, Z.**, **Zhang, L.**, **Wang, J.** & **Wu, C.** (2020) Glycosyltransferase GT1 family: phylogenetic distribution, substrates coverage, and representative structural features. *Computational and Structural Biotechnology Journal*, **18**, 1383–1390.

Zhang, Z., **Zou, W.**, **Lin, P.**, **Wang, Z.**, **Chen, Y.**, **Yang, X.** ***et al.*** (2024) Evolution and function of MADS-box transcription factors in plants. *International Journal of Molecular Sciences*, **25**, 13278.

Zheng, X., **Chen, D.**, **Chen, B.**, **Liang, L.**, **Huang, Z.**, **Fan, W.** ***et al.*** (2021) Insights into salvianolic acid B biosynthesis from chromosome-scale assembly of the *Salvia bowleyana* genome. *Journal of Integrative Plant Biology*, **63**, 1309–1323.

Zheng, Y., **Jiao, C.**, **Sun, H.**, **Rosli, H.G.**, **Pombo, M.A.**, **Zhang, P.** ***et al.*** (2016) iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Molecular Plant*, **9**, 1667–1670.