Plant Communications
Research article

CellPress
Partner Journal

The genome of giant waterlily provides insights into the origin of angiosperms, leaf gigantism, and stamen function innovation

Xiaohui Wen[1,2,14], Yuwei Liang[1,14], Hongyan Shan[3,14], Xiaojun Chang[1,14], Xiaoming Song[4,14], Shaoqin Shen[4], Yanhong Fu[4], Dan Chen[5], Fei Chen[6], Yueqing Li[7], Qian Guan[7], Qiang Gao[1], Qi Wang[1], Yonglin Li[8], Zhengjia Wang[9], Hongzhi Kong[3], Huan Liu[10,*], Xiang Gao[7,*], Xiaofan Zhou[8,*], Chris Thorogood[11,12,*] and Liangsheng Zhang[1,13,*]

[1]Zhejiang Key Laboratory of Horticultural Crop Quality Improvement, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, China

[2]Zhejiang Institute of Landscape Plants and Flowers, Zhejiang Academy of Agricultural Sciences, Hangzhou, China

[3]State Key Laboratory of Plant Diversity and Specialty Crops, Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, China

[4]College of Life Sciences, North China University of Science and Technology, Tangshan, Hebei 063210, China

[5]Hainan Institute of Zhejiang University, Sanya, China

[6]Hainan Key Laboratory for Sustainable Utilization of Tropical Bioresources, College of Tropical Crops, Hainan University, Haikou, China

[7]Key Laboratory of Molecular Epigenetics of MOE and Institute of Genetics & Cytology, Northeast Normal University, Changchun, China

[8]Guangdong Laboratory for Lingnan Modern Agriculture, Guangdong Province Key Laboratory of Microbial Signals and Disease Control, Integrative Microbiology Research Center, South China Agricultural University, Guangzhou, China

[9]National Key Laboratory for Development and Utilization of Forest Food Resources, Zhejiang A&F University, Hangzhou, Zhejiang, China

[10]Key Laboratory of Genomics, Ministry of Agriculture, BGI Research, Shenzhen, China

[11]Department of Biology, University of Oxford, Oxford OX1 3SZ, UK

[12]University of Oxford Botanic Garden and Arboretum, Oxford OX1 4AZ, UK

[13]Yazhouwan National Laboratory, Sanya, China

[14]These authors contributed equally

***Correspondence: Huan Liu (liuhuan@genomics.cn), Xiang Gao (gaoxiang424@163.com), Xiaofan Zhou (xiaofan_zhou@scau.edu.cn), Chris Thorogood (chris.thorogood@obg.ox.ac.uk), Liangsheng Zhang (zls83@zju.edu.cn)**

https://doi.org/10.1016/j.xplc.2025.101342

ABSTRACT

As some of the earliest evolving flowering plants, waterlilies offer unique insights into angiosperm evolution. Giant Amazonian waterlilies (genus *Victoria*) are of particular interest due to their production of the world's largest floating leaves and gigantic flowers that entrap pollinating beetles. Here, we report chromosome-level genome assemblies of *Victoria cruziana* and three related waterlilies: *Euryale ferox*, *Nymphaea mexicana*, and *Brasenia schreberi*. We found an ancient whole-genome duplication event specific to the Nymphaeales. We reveal major gene duplication and loss events throughout the evolution of angiosperms, with substantial implications for flower development and the biosynthesis of floral volatile organic compounds (FVOCs) in waterlilies. Importantly, we report a unique division of labor in the stamen function of *V. cruziana* linked to beetle attraction by FVOCs. This is related to the ultra-high expression of *VicSABATHa* along with *Vicchitinase*, possibly linked to protection from damage by trapped beetles. Overexpression of *VicSABATHa* in tobacco leaves reveals a capacity to produce volatile fatty acids, confirming its role in their catalytic synthesis. Overall, these findings provide novel insights into the evolution and adaptations of waterlilies and flowering plants in general.

Key words:: waterlily genomes, early angiosperms, leaf gigantism, stamen innovation, floral scent biosynthesis

Wen X., Liang Y., Shan H., Chang X., Song X., Shen S., Fu Y., Chen D., Chen F., Li Y., Guan Q., Gao Q., Wang Q., Li Y., Wang Z., Kong H., Liu H., Gao X., Zhou X., Thorogood C., and Zhang L. (2025). The genome of giant waterlily provides insights into the origin of angiosperms, leaf gigantism, and stamen function innovation. Plant Comm. **6**, 101342.

INTRODUCTION

Amborellales, Nymphaeales, and Austrobaileyales represent the earliest diverging lineages of flowering plants (Friis et al., 2018; Povilus et al., 2020; Yang et al., 2020; Zhang et al., 2020). These groups hold the key to unraveling "Darwin's abominable mystery," which refers to the rapid diversification of angiosperms after their initial emergence approximately 150 million years ago (Doyle, 2012; Chen et al., 2017a; 2017b; Buggs, 2017). The order Nymphaeales comprises three families: Cabombaceae, Hydatellaceae, and Nymphaeaceae (Angiosperm Phylogeny Group IV system); research on these plants has provided valuable insights into the broader evolution of angiosperms (Zhang et al., 2020).

Giant waterlilies (genus *Victoria*, order Nymphaeales) are aquatic plants of substantial horticultural and economic importance. *Victoria cruziana* produces gigantic leaves (>2.5 m across) and enormous flowers (40 cm across), forming vast networks on the surface of the Parana river basin and its tributaries (Smith et al., 2022). The giant leaves provide a notable example of biomimetics—the design of systems and materials inspired by nature—in architecture, famously influencing the support structure for the large glass panes of the Victorian Crystal Palace in London (Brebbia and Assoc, 2007). The plant achieves these proportions by producing a leaf that is structurally more efficient than those of smaller waterlily species; the varying thickness and regular branching of its veins ensure sufficient structural integrity to allow extensive coverage of the water surface, enabling optimal light capture despite relatively low leaf biomass (Box et al., 2022). Despite the esthetic appeal and scale of this botanical enigma, little is known about the molecular mechanisms that govern the mechanics of these extraordinary leaves.

The genus *Victoria* is also notable for its unusual flowers and pollination system. Pollination in *Victoria* has been known to involve *Cyclocephala* beetles since Robert Schomburgk's report (Ervik and Knudsen, 2003; Thien et al., 2009). The flowers bloom over two consecutive nights in a progressive sequence. Upon opening, the giant flowers are white and emit a strong, fruity odor that attracts night-flying beetles, which fertilize the receptive pistils (Ervik and Knudsen, 2003). At this stage, the stamens are erect and form a tunnel into the flower; the petals remain erect and white throughout the night and into the following morning. The innermost whorls of stamens (and inner staminodia) then close inward to form a floral chamber, within which the pollinating beetles remain captive as they feed and mate (Ervik and Knudsen, 2003; Endress, 2010; Jana, 2018). During this period, the temperature of the floral chamber rises to 7°C–10°C above ambient air temperature (Schimpf et al., 2017). By the evening of the first day and throughout the second day, the stamens ripen and release their pollen (Schimpf et al., 2017). This system of dichogamy, in which female parts of the flower mature before the male parts, ensures cross-fertilization. Moreover, the flowers exhibit thermogenesis, a pattern of floral heating associated with beetle attraction and considered an energy reward (Bernhardt, 2000; Seymour and Matthews, 2006). Large flower size may be an evolutionary consequence of selection for mass congregation of pollinating beetles, thereby maximizing the probability of pollination (Thien et al., 2009). Understanding the molecular mechanisms that regulate the pollination biology of *Victoria* is important because the plant is widely celebrated and cultivated in botanical gardens worldwide. Despite reports of seed production after artificial selfing, few published studies exist regarding controlled pollination in the genus (Smith et al., 2022). Propagation remains hampered by poor seed set, and further work is required to understand the mating system of these enigmatic plants.

In the context of a sixth mass extinction (the Anthropocene), genomes of non-model species are becoming an increasingly important tool in conservation biology (Paez et al., 2022). Few waterlily species have undergone genome sequencing; notable examples include *Nymphaea colorata*, *Nymphaea thermarum*, and *Euryale ferox* (Povilus et al., 2020; Yang et al., 2020). Here, we report chromosome-level genomes of *V. cruziana* (2n = 24, Nymphaeaceae), *E. ferox* (2n = 58, Nymphaeaceae), *Nymphaea mexicana* (2n = 54, Nymphaeaceae), and *Brasenia schreberi* (2n = 72, Cabombaceae) to enrich genomic data regarding early angiosperms and to investigate genome evolution history in the Nymphaeales. Further comparative genomic and transcriptomic analyses will enable us to (1) enhance our understanding of angiosperm evolution; (2) decipher the molecular mechanisms underlying development, with an emphasis on leaf gigantism, flower development, and floral volatile organic compound (FVOC) biosynthesis, particularly in relation to their adaptive implications; and (3) provide genomes of non-model species in a group of plants particularly vulnerable during this period of unprecedented global change.

RESULTS

The genome assemblies of four waterlily species

We assembled chromosome-level genomes for *V. cruziana* (3.26 Gb), *E. ferox* (0.73 Gb), *N. mexicana* (0.50 Gb) (Nymphaeaceae), and *B. schreberi* (1.2 Gb) (Cabombaceae) (Figure 1 and Supplemental Figure 1; Supplemental Tables 1 and 2). For the first three species, we obtained chromosome-level genomes based on high-throughput chromosome conformation capture (Hi-C) data. For *B. schreberi*, we anchored our assembly to the chromosomal level based on the recently published genome. Benchmarking Universal Single-Copy Orthologs (BUSCO) results showed 90.84%–94.84% completeness across the six Nymphaeales genomes (Figure 1C). We found that repetitive elements constituted 91.43% of the *V. cruziana* genome, of which 79.72% were long terminal repeat (LTR) retrotransposons (Supplemental Tables 1 and 2; Supplemental Figure 2). *Copia* and *Gypsy* elements comprised the majority of LTRs in *V. cruziana*; *Gypsy elements constituted* 75.91% (Supplemental Table 3; Supplemental Figure 3). The proportion of repetitive elements in the other three sequenced waterlilies ranged from approximately 39.07% to 52.06% (Supplemental Table 3; Supplemental Figure 3).

Frequent polyploidization in waterlilies

Collinearity and *Ks* distributions reveal an ancient WGD event (peak 1) shared by Nymphaeales during the evolution of waterlilies (Figures 2A and 2C and Supplemental Figures 3–5). Peak 4, representing the divergence of *V. cruziana* and *B. schreberi*, after

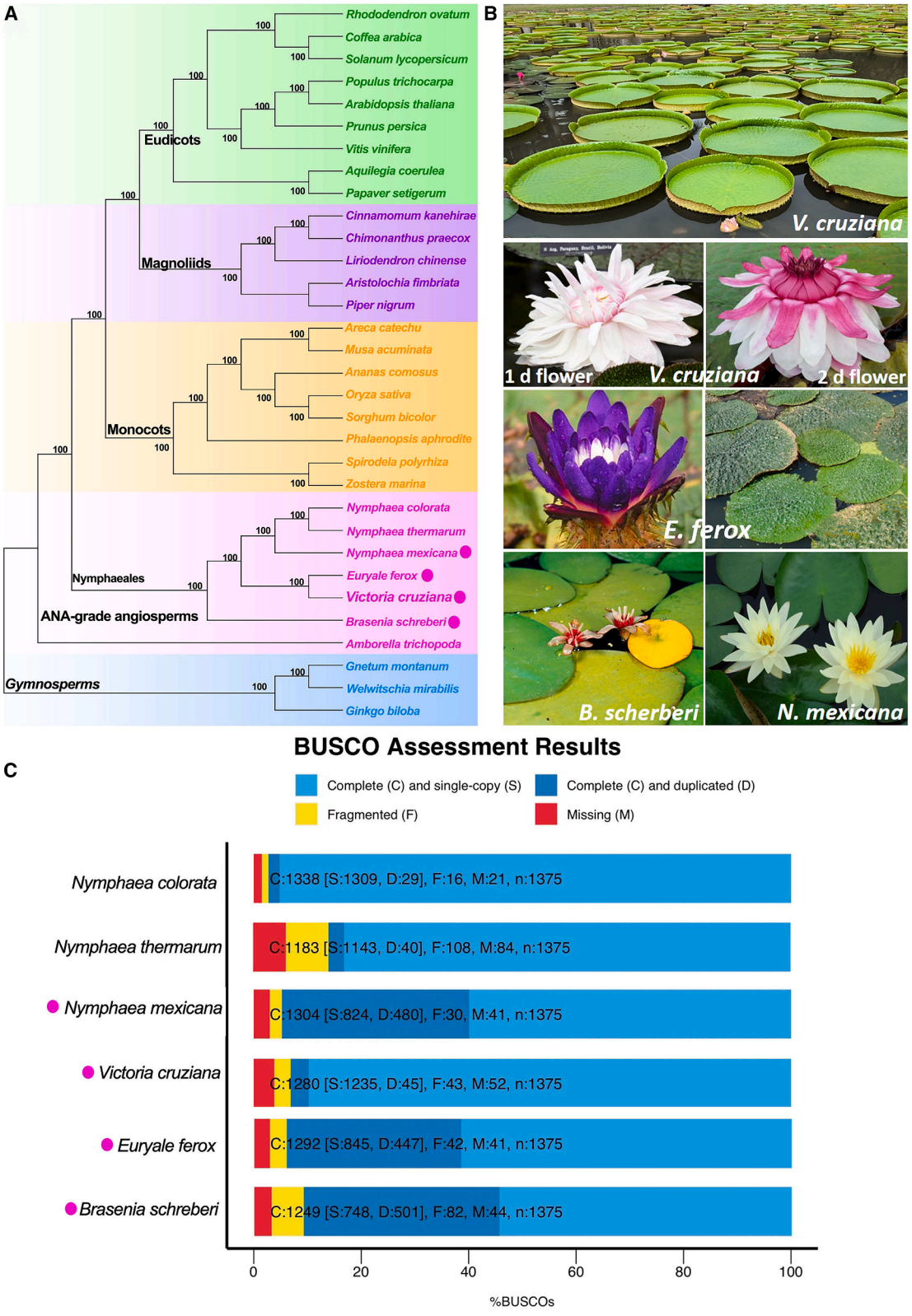

Figure 1. Phylogenetic relationships of waterlilies with other plants and genome assemblies of *V. cruziana*, *E. ferox*, *N. mexicana*, and *B. schreberi* in the present study.

(A) Phylogeny of 29 representative angiosperms, including six waterlily species: *V. cruziana*, *E. ferox*, *N. mexicana*, *B. schreberi*, *N. thermarum*, and *N. colorata*.

(B) Morphology of leaves and flowers in *V. cruziana*, *E. ferox*, *B. schreberi*, and *N. mexicana*.

(C) BUSCO assessment of six waterlily genomes.

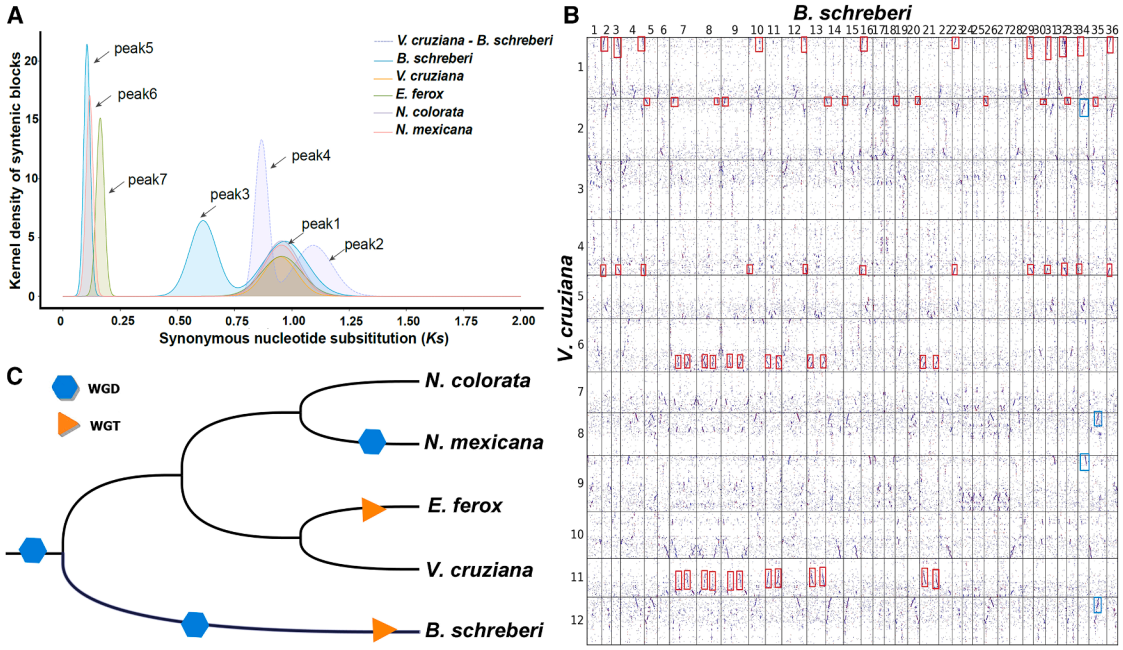

Figure 2. WGT and WGD events in Nymphaeales.

(A) Density distribution of estimated synonymous substitution rates (*Ks*) of syntelog pairs from intragenomic comparisons of *V. cruziana*, *B. schreberi*, *E. ferox*, *N. mexicana*, and *N. colorata*, and an intergenomic comparison between *V. cruziana* and *B. schreberi*. Labeled peaks indicate potential WGD events.

(B) Syntenic dot plots for *V. cruziana* and *B. schreberi*. Red rectangles represent collinear blocks in *B. schreberi*; blue rectangles represent collinear blocks in *V. cruziana*.

(C) Phylogenetic framework and genome evolution history, including WGTs and WGDs, in waterlilies. WGD, whole genome duplication; WGT, whole genome triplication.

the peaks of other whole-genome duplication (WGD) events in Nymphaealestherefore, these WGD events occurred before the divergence of *V. cruziana* and *B. schreberi* (Figure 2A). Following this divergence (peak 4), the *B. schreberi* genome underwent two WGD events (peaks 3 and 5) (Figure 2A). The distribution of these peaks is also supported by collinearity plots. According to the orthology ratio of 2:12 between the *V. cruziana* and *B. schreberi* genomes in the homologous gene dot plot, *B. schreberi* experienced two WGD events (peak 1 and peak 3) and one whole-genome triplication (WGT) event (peak 5). (Figure 2A and 2B). The *Ks* distributions of *V. cruziana* (2n = 24) and *B. schreberi* (2n = 72) indicate that *B. schreberi*, belonging to Cabombaceae, shares the ancient WGD event identified in Nymphaeaceae (peak 1) (Figure 2A and 2B). Compared with the genomes of *N. colorata* and *V. cruziana*, the *N. mexicana* genome has undergone diploidization (peak 6), whereas that of *E. ferox* has experienced triploidization (peak 7) (Figure 2 and Supplemental Figures 4 and 5).

Gene duplication and loss events in Nymphaeales

Gene duplication and loss events are considered key evolutionary forces driving the adaptive phenotypic diversity of angiosperms (Bowers et al., 2003; Jiao et al., 2011; Liu et al., 2021). However, genomic data for angiosperms are not comprehensive due to the limited availability of genome sequences from Amborellales, Nymphaeales, and Austrobaileyales. We generated genomic sequence data for four Nymphaeales species and combined them with two published genomes, *N. colorata* and *N. thermarum*, to identify 18 606 ortholog clusters across 29 angiosperm genomes, encompassing four lineages: eudicots, monocots, magnoliids, and Nymphaeales (Figure 3 and Supplemental Figure 6). Approximately 59.35% of ortholog clusters are shared across these lineages (Figure 3A), including *AOX1* (Supplemental Figure 7), *UCP1/2/3* (Supplemental Figures 8 and 9), *AP3* (Supplemental Figure 10), and a defense-related ortholog cluster (*ESD1*) (Supplemental Figure 11), indicating that these genes may have existed in early angiosperms and were preserved during angiosperm evolution. Some of these genes have retained their conserved roles in plant growth and development. For example, the *AP3* ortholog cluster across the four lineages consistently participates in stamen and petal identity (Ng and Yanofsky, 2001). In contrast, other genes have evolved diverse functions related to adaptation. For example, the *AOX1* and *UCP* ortholog clusters, which regulate the flower chamber and thermogenesis in Nymphaeales, have shifted to heat resistance roles in eudicots (Wagner et al., 2008; Keinan et al., 2021).

Gene duplication, along with neofunctionalization and subfunctionalization, is mainly driven by polyploidization during angiosperm evolution (Albalat and Canestro, 2016). Our study identified gene duplication events in Nymphaeales, such as the ortholog cluster of the defense-related gene *NRG1* (Supplemental Figure 12). We also found that the ortholog involved in fatty acid methyl ester synthesis—salicylic acid, benzoic acid, and theobromine synthase (*SABATH*)—underwent gene duplication events in Nymphaeales, which might be related

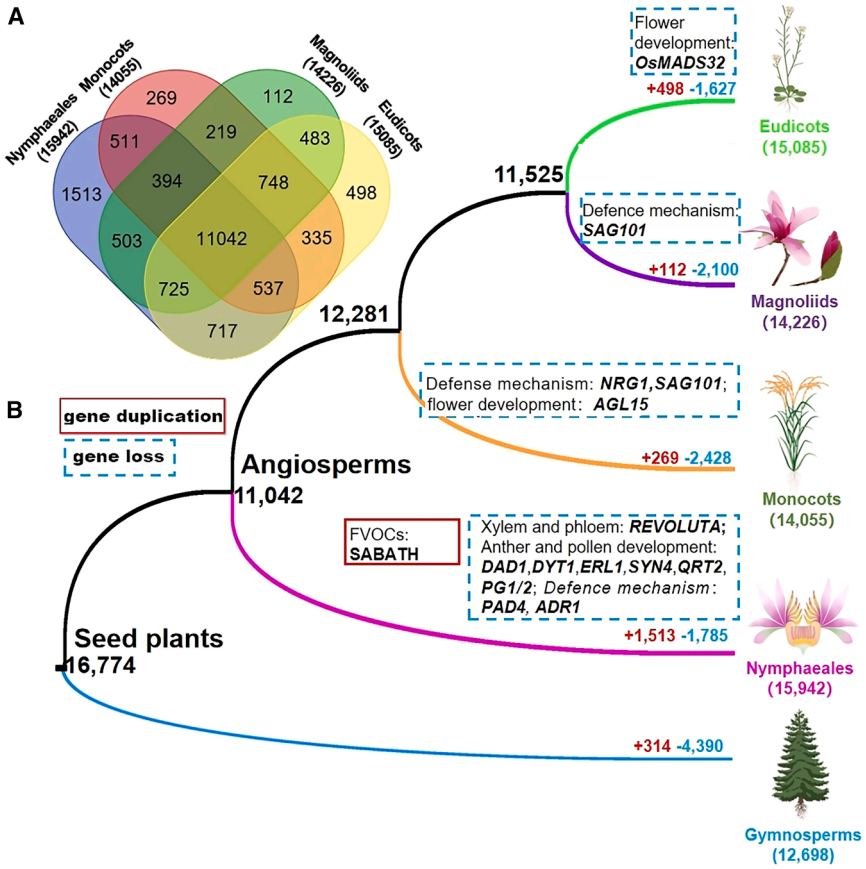

Figure 3. Gene duplication and loss events in the evolutionary history of angiosperms.
(A) Venn diagram showing the numbers of specific and common gene families in Nymphaeales, monocots, magnoliids, and eudicots.
(B) Simplified phylogenetic relationships of major angiosperm clades, highlighting gene family number variation and key genes involved in molecular regulatory mechanisms (Supplemental Figures 7–25). Genes in solid red boxes represent genes gained in the lineage; genes in dashed blue boxes represent genes lost in the lineage.

to unique floral scent biosynthesis in waterlilies (Supplemental Figure 13) (Zhang et al., 2020). Notably, we detected a substantial number of gene loss events, which are considered an adaptive evolutionary force, particularly effective during abrupt environmental challenges (Albalat and Canestro, 2016). Our study supports previous findings regarding gene loss events based on the genomes of early-diverging angiosperms. For example, ortholog clusters of the flower development-related gene *AGL15* (Supplemental Figure 14) and resistance-related genes *TNL*, *NRG1*, and *SAG101* (Supplemental Figures 12, 15, and 16) are all lost in representative monocots (Becker and Günter, 2003; Chen et al., 2017a; 2017b; Liu et al., 2021). Additionally, the flower development-related gene *OsMADS32* has been lost in eudicots (Supplemental Figure 17) (Chen et al., 2017a; 2017b). In Nymphaeales, we identified 1785 gene loss events, including resistance-related genes *PAD4* and *ADR1* (Supplemental Figures 18 and 19) and a xylem and phloem development gene *REVOLUTA* (Supplemental Figure 20). These genes are associated with resistance mechanisms and xylem degeneration, contributing to the adaptation of Nymphaeales to aquatic environments (Povilus et al., 2020). Gene loss events in ortholog clusters related to pollen development (*DYT1*, *SYN4*, and *ERL1*) (Supplemental Figures 21–23) and anther dehiscence (*DAD1*, *QRT2*, and *PG1/2*) (Supplemental Figures 24 and 25 and Figure 3) were also identified. These gene loss events might be involved in protandry during flower development in Nymphaeales. In summary, by enriching the available waterlily genomes, we provide further insights into gene duplication and loss events that may regulate plant traits potentially associated with specific adaptations to aquatic environments during angiosperm evolution.

Leaf gigantism in *V. cruziana*

We performed weighted gene co-expression network analysis (WGCNA) to identify genes potentially responsible for leaf gigantism in *V. cruziana* (Supplemental Figures 26–49). In the early stages of leaf development (leaves 0.5 and 2 cm in diameter, respectively), over 60% of genes co-expressed in the turquoise module were enriched in cell proliferation processes, suggesting that cell proliferation plays a critical role in early leaf development and size determination (Figure 4A and Supplemental Figures 29 and 30). Multiple genes in the turquoise module demonstrated involvement in controlling leaf or other organ size via the regulation of cell proliferation, brassinosteroid (BR) signaling, and BR biosynthesis. These genes included orthologs of *GRFs*, *GIF1*, *SPLs*, *CYCD3;1*, *UBP15*, *ER*, *ERL1* and *ERL2*, *CYP51G1s*, *CPD*, *BRI1*, and *BZR1* (Figure 4B and Supplemental Figures 31–34). Among them, orthologs of *CYCD3;1*, *ER*, *SPLs*, *CYP51G1s*, *CPD*, *BRI1*, and *BZR1* were co-expressed in the WGCNA-generated turquoise module correlated with early leaf development in *V. cruziana* (Supplemental Figure 35). *V. cruziana* not only produces gigantic leaves but also bears large flowers (~40 cm in diameter), significantly exceeding the size of most angiosperm flowers. Considering that plant growth and development processes are often regulated by a series of conserved genetic mechanisms, the genes involved in leaf size development may exhibit similar expression patterns during flower size development. Similarly, the genes discussed above showed high expression levels in the early stages of petal development (petals 1 and 2 cm in length, respectively) (Figure 4A and 4B). In modules correlated with the cell expansion phase in *V. cruziana* (i.e., magenta, yellow, and green modules), genes involved in cell wall modification contributing to cell expansion were significantly expressed in leaves during mid-development. These genes included orthologs of *XTHs*, *PMEIs*, *EXPs*, and others (Figure 4A and Supplemental Figure 36). During the final stages of leaf development, the lamina unfurls and the vertical leaf margin elevates—a feature that strengthens the mature leaf. This stage was correlated with genes enriched in the brown and purple modules (Supplemental Figure 29). Expanded genes related to

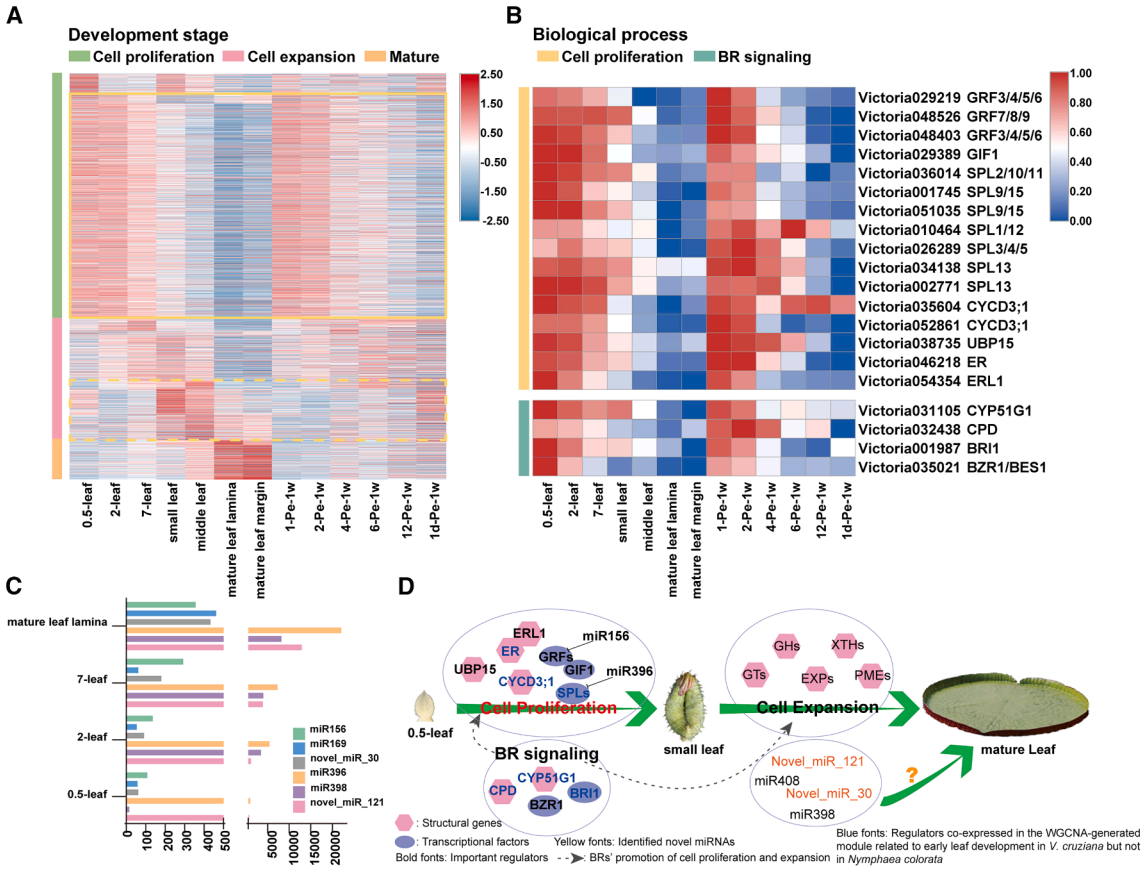

Figure 4. Developmental mechanisms underlying leaf gigantism.

(A) Heatmap showing relative expression profiles of genes in 12 co-expression modules in*V. cruziana* leaves and petals, generated by WGCNA. The 12 modules were divided into three groups corresponding to three stages of leaf development. Genes in solid boxes in Figure 2A are correlated with the turquoise module shown in Supplemental Figure 24. Genes in dashed boxes correspond to the magenta, yellow, and green modules in Supplemental Figure 24.

(B) Heatmap showing the expression profiles of key genes involved in early-stage leaf and petal growth.

(C) Histogram showing the expression profiles of putative functional miRNAs involved in leaf size regulation in *V. cruziana*.

(D) Proposed molecular mechanism underlying leaf gigantism in *V. cruziana*.

energy production—such as those involved in redox reactions, fatty acid metabolism, and glycometabolism—may support rapid leaf growth and the development of gigantic flowers (Supplemental Figure 37).

Analysis of microRNA (miRNA) expression patterns during leaf development (S1, S2, S3, and S4, where S1 corresponds to early leaf development, and S2 and S3 correspond to middle and late stages, respectively) in *V. cruziana* identified multiple miRNAs that potentially contribute to organ size regulation (Supplemental Figures 38 and 39; Supplemental Table 4) (Rodriguez et al., 2010; Yang et al., 2021). Notably, miR396, miR398, and novel_miR_121 may play roles in leaf development (Figure 4C) (Alemán-Báez et al., 2024). In particular, miR396 exhibited very high expression levels, with a fragments per kilobase of transcript per million fragments (FPKM) value reaching 20 000 in S4—almost a 50-fold increase relative to miR169 (Figure 4C). Based on these findings, we propose that cell proliferation-related genes (e.g., *CYCD3;1*), BR signaling process-related genes (e.g., *SPLs* and *GRFs*), and miRNAs (e.g., miR396 and miR156) play key roles in regulating leaf gigantism in *V. cruziana* (Figure 4D and Supplemental Figure 40).

Floral organ identity, dichogamy, and stamen differentiation of *V. cruziana*

To gain insights into the floral development of *V. cruziana*, we examined the expression profiles of putative A-, B-, C-, D-, and E-function genes across different types of floral organs at various developmental stages (Supplemental Figures 27, S28, and 41–43). Putative B-function (*VicAP3* and *VicPI*) and E-function (*VicSEP*) genes were broadly expressed across floral organs (Figure 5A and Supplemental Figure 41). In contrast, expression of the putative A-function gene *VicAGL6* gradually decreased from outer to inner whorls, whereas expression of the C-function genes *VicAGa* and *VicAGb* gradually increased from outer to inner whorls (Figure 5A). Dichogamy, the temporal separation of stamens and carpels, ensures cross-pollination in angiosperms (Ervik and Knudsen, 2003; Sargent and Otto, 2004). The dichogamy type of *V. cruziana* is metandry; the carpels mature on the first day of blooming, followed by pollen

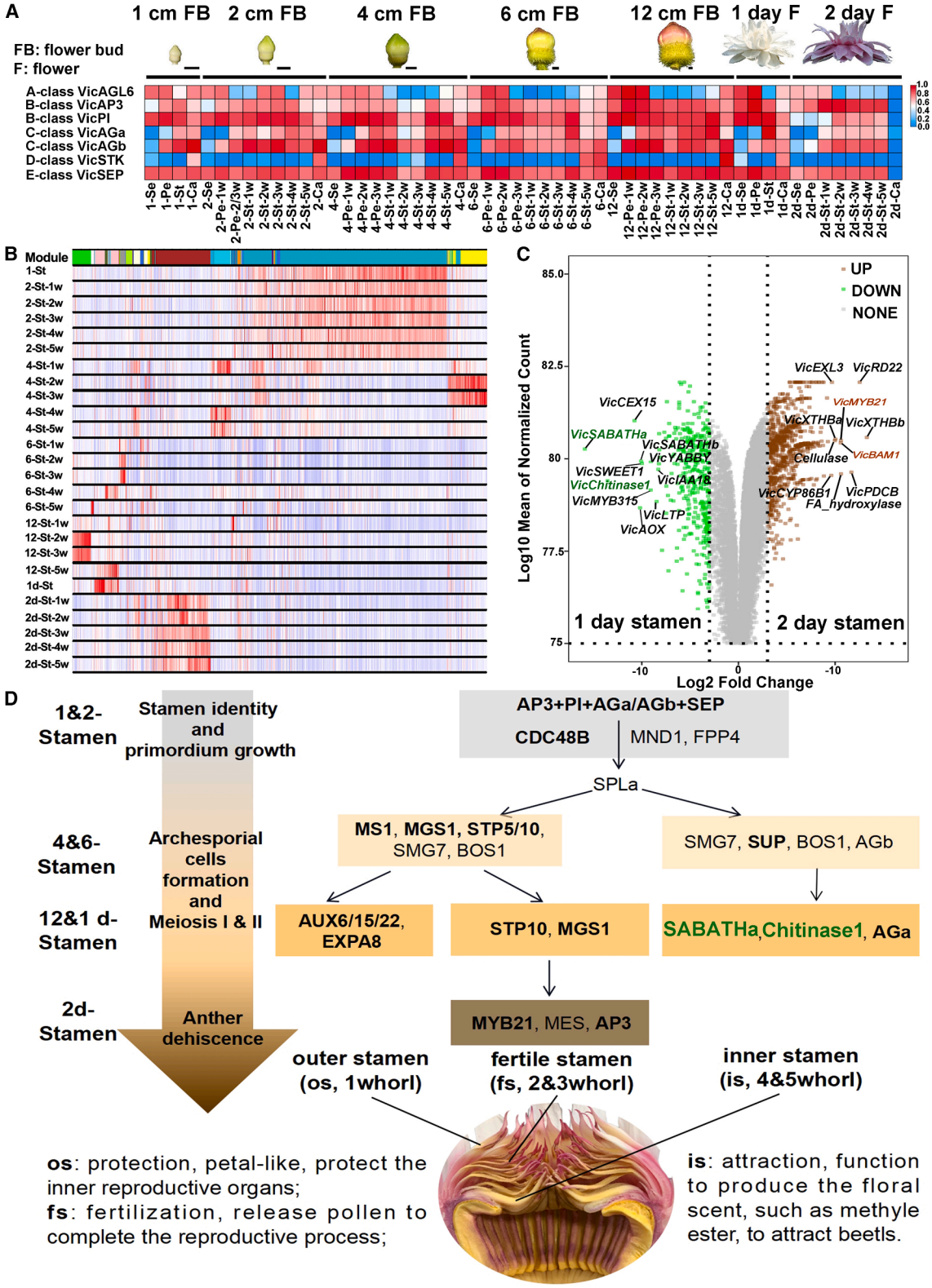

Figure 5. Molecular regulatory mechanisms of flower development, with emphasis on stamen development, in *V. cruziana*.
(A) Heatmap of ABC(D)E gene expression. Scale bars: 1 mm. Se, sepal; Pe, petal; St, stamen; Ca, carpel; w, whorl.
(B) Heatmap showing gene expression in 16 co-expression modules across various floral whorls and stamen developmental stages.
(C) Volcano plot showing differentially expressed genes between stamens from day 1 and day 2 flowers.
(D) Key regulatory genes involved in the development of different stamen whorls in *V. cruziana*.

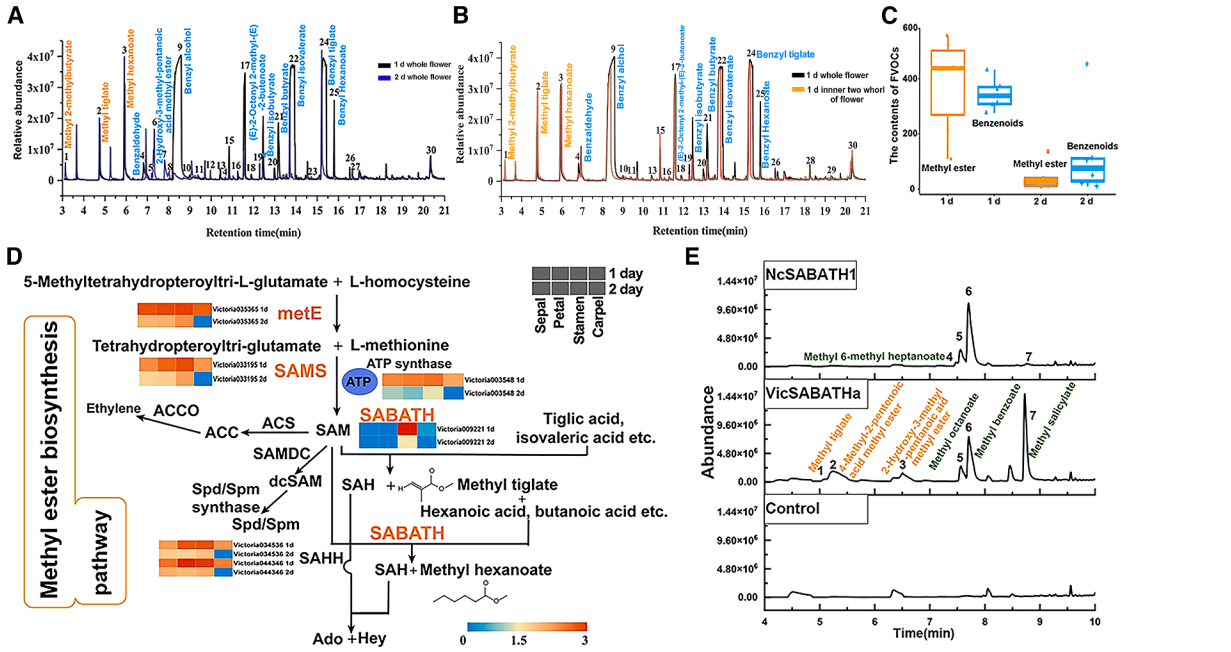

Figure 6. Floral Volatile Organic Compounds (FVOCs) and pollination syndrome in *V. cruziana*.

(A) GC-MS peak plot showing FVOCs emitted from whole flowers of *V. cruziana* during 2 consecutive days of blooming. The x-axis represents retention time; the y-axis represents relative abundance.

(B) GC-MS peak plot of volatile compounds emitted from the inner two whorls of *V. cruziana* (mainly comprising stamens). The x-axis represents retention time; the y-axis represents relative abundance.

(C) GC-MS peak plot showing emissions of methyl esters and benzenoids from *V. cruziana* flowers during 2 consecutive days of blooming. The x-axis represents retention time; the y-axis represents relative abundance.

(D) Methyl ester biosynthesis pathway in *V. cruziana*. Ado, adenosine; Hcy, homocysteine. Genes expressed on the first and second blooming days are displayed in rows.

(E) Functional validation of *VicSABATHa and NcSABATH1 through GC-MS analysis of volatiles in N. benthamiana* overexpression lines. The x-axis represents retention time; the y-axis represents relative abundance.

release on the second day. Carpel maturation might be modulated by the molecular regulation of C-, D-, and E-function genes. Based on phenotypic observations and WGCNA results across various stamen developmental stages, *V. cruziana* flowers possess three types of stamens: outer petaloid stamens (represented by first whorl stamens in flower buds with diameters of 4-, 6-, and 12-cm; 4&6&12-St-1w), middle fertile stamens (represented by second- and third-whorl stamens in flower buds with diameters of 4-, 6-, and 12-cm, respectively; 4&6&12-St-2∼3w), and inner staminodes (represented by fourth and fifth-whorl stamens in flower buds with diameters of 4-, 6-, and 12-cm; 4&6&12-St-4∼5w) (Figure 5). WGCNA results revealed that the stamen primordia in flower buds with diameters of 1- (1-St) and 2-cm (2-St-1∼5w) do not undergo differentiation at those stages (Figure 5B and Supplemental Figures 44–46). Stamen primordia gradually differentiate into inner petaloid, fertile, and sterile stamens beginning in flower buds with diameters of 4-cm and continuing until anthesis (Figure 5B and Supplemental Figures 47–54; Supplemental Tables 5–7). The development of inner stamens was regulated by pink-module genes, along with the hub-regulated genes *VicAGb*, *VicSABATHa*, and *Vicchitinase1*, during the first day of blooming (Figure 5C and 5D; Supplemental Table 8; Supplemental Figures 51 and 52). In contrast, the expression of *VicSTP10* and *VicMGS1* in fertile stamens (12-St-2∼3w) may be associated with the development of fertile stamens, particularly the processes of anther and pollen development during the first day of blooming (Figure 5C–5E and Supplemental Figures 51 and 52). By the second day, the expression of *VicAP3*, *VicMYB21*, and *VicMES* in the brown module may regulate anther dehiscence, contributing to dichogamy (Figure 5C–5E and Supplemental Figures 53–56). Based on the fading border model used to identify the four types of floral organs (sepals, petals, stamens, and carpels), different gene clusters regulate the differentiation of stamens into inner, fertile, and outer subtypes (Figure 6G).

The biosynthesis of floral volatile organic compounds (FVOCs) in *V. cruziana*

We investigated the biosynthesis of *V. cruziana* FVOCs over 2 consecutive days of blooming (Figure 6A–6C). On the first day, 52.45% of the FVOCs were methyl esters (Figure 6C; Supplemental Tables 9 and 10). The quantities of the two principal FVOCs substantially decreased over the course of blooming, especially the content of methyl esters, which constituted only 25.41% on the second day (Figure 6C; Supplemental Tables 9 and 10). The methyl esters in *V. cruziana* were mainly composed of methyl 2-methylbutyrate, methyl tiglate, and methyl hexanoate (Figure 6C; Supplemental Tables 9 and 10). In contrast, the FVOCs of *N. colorata* primarily included fatty acid derivatives—representing 95% and 87% on

the first and second blooming days, respectively. These fatty acid derivatives were predominantly long-chain fatty acid derivatives (Supplemental Figure 57).

Previous studies have suggested that the SABATH family of methyltransferase 19 enzymes is involved in methyl ester biosynthesis in waterlilies (Zhao et al., 2008; Zhang et al., 2020; Jiang et al., 2021). In accordance with this hypothesis, our study focused on *VicSABATHa*. We identified the top 20 most highly expressed genes in the inner stamens (represented by 12-St-5w tissue) (Supplemental Figures 58–60). Intriguingly, the gene with the highest expression in 12-St-5w was *VicSABATHa*, which encodes the SABATH enzyme, followed by *Vicchitinase1*, which is involved in hydrolyzing glycosidic bonds in chitin for pathogen resistance (Grover, 2012) (Supplemental Figures 58–60). WGCNA also showed that these two genes were enriched in the pink module, which was associated with inner stamen (12-St-5w) development (Figure 5B; Supplemental Figure 58). The top 20 most highly expressed genes in the 12-St-5w of *V. cruziana* also include genes related to methyl ester synthesis (*VicSAMS*, *VicSAHH*, and *VicmetE*), phenylpropanoid volatile organic compound synthesis (*VicADH* and *VicAcyl*-coenzyme A-binding protein), and defense response (*VicDefensin* and *VicnsLTP*) (Supplemental Figures 59–61). Based on the reported floral scent biosynthesis pathway, orthologs of *VicSAMS*, *VicSAHH*, and *VicmetE* act upstream of the ortholog of *SABATHa* in the methyl ester biosynthesis pathway (Dudareva et al., 2013). We present a probable methyl ester biosynthesis pathway for *Victoria* (Figure 6D and Supplemental Figure 62). Genes involved in the methyl ester synthesis pathway (*VicmetE*, *VicSAMS*, *VicSAHH*, *VicATP* synthase, and *VicSABATHa*) showed consistently high levels of expression in stamen tissues on the first day of blooming (Figure 6D and Supplemental Figure 62). VicSABATHa functions as a key enzyme in the final step of the methyl ester synthesis pathway (Figure 6D). Taken together, these findings suggest that *VicSABATHa* is a key gene involved in methyl ester biosynthesis in *V. cruziana*.

Furthermore, we examined the biochemistry of VicSABATHa and NcSABATH1. *In vitro* methyltransferase enzyme assays revealed that the recombinant VicSABATHa protein exhibits high catalytic activity concerning the formation of short-chain fatty acids, including butyric, tiglic, and hexanoic acids—essential compounds in methyl ester biosynthesis (Supplemental Figure 63). In contrast, the recombinant NcSABATH1 protein preferentially catalyzes the formation of long-chain fatty acids, such as decanoic acid (Supplemental Figure 63). We also performed *in vivo* enzyme activity verification experiments using transiently expressed *VicSABATHa* and *NcSABATH1* in tobacco leaves (Figure 6E). Tobacco leaves overexpressing *VicSABATHa* released a greater variety of short-chain fatty acid methyl esters, including methyl tiglate, 4-methyl-2-pentenoic acid methyl ester, and 2-hydroxy-3-methyl-pentanoic acid methyl ester (Figure 6E). Overexpression of *NcSABATH1* in tobacco leaves produced methyl 6-methyl heptanoate, methyl octanoate, methyl benzoate, and similar compounds (Figure 6E). Proteins extracted from tobacco leaves overexpressing *VicSABATHa* catalyzed butyric, tiglic, hexanoic, and decanoic acids (Supplemental Figure 64). Conversely, proteins extracted from tobacco leaves overexpressing *NcSABATH1* only catalyzed decanoic acid (Supplemental Figure 64). Furthermore, enzyme kinetic analysis revealed the highest catalytic efficiency of VicSABATHa for tiglic acid, followed by hexanoic acid; the lowest efficiency was observed for decanoic acid (Supplemental Figure 65; Supplemental Table 11). The catalytic efficiency of NcSABATH1 for capric acid was similar to that of VicSABATHa for tiglic acid. However, due to the extremely low catalytic ability of NcSABATH1 toward tiglic and hexanoic acids, reliable kinetic parameters could not be obtained. Overall, these results suggest that *VicSABATHa* is specifically expressed in the inner stamens of *V. cruziana*, where it contributes to short-chain fatty acid methyl ester biosynthesis; *NcSABATH1* is highly expressed in petals and is responsible for the biosynthesis of medium-chain fatty acid methyl esters in *N. colorata*. Our findings indicate that the *VicSABATHa* gene serves as a pivotal catalytic gene that has undergone further evolution in *V. cruziana*.

DISCUSSION

Evolution of waterlilies and other angiosperms

Angiosperms represent one of the most species-rich clades in the tree of life and one of the most spectacular terrestrial radiations on Earth. Waterlilies, in particular, offer valuable insights into angiosperm evolution—which remains enigmatic in many respects—as well as their adaptability to diverse environments (Yang et al., 2020; Zhang et al., 2020). Two genomes of *E. ferox* and the genome of *B. schreberi* have previously been published (Yang et al., 2020; Wu et al., 2022; Lu et al., 2023). One of the *E. ferox* genomes highlights its strong environmental adaptability. We also report the genomes of *E. ferox* and *B. schreberi*; however, unlike previous studies, our analysis of *B. schreberi* includes a detailed assessment of its WGD events in relation to other waterlilies. We identified an ancient WGD event during the evolution of waterlilies, as well as several recent WGDs shared within Nymphaeaceae. We anchored our *B. schreberi* genome to the chromosome level based on the previously published *B. schreberi* genome (Lu et al., 2023). Based on our chromosome-level assembly, we found that *B. schreberi* shares the ancient WGD event observed in Nymphaeaceae. Moreover, we show that *B. schreberi* underwent two WGD events and one WGT event. These findings suggest that waterlilies have undergone frequent polyploidization events.

We also report the genome of *N. mexicana*, a temperate waterlily species. Comparative genomic analysis of this species with the subtropical and tropical waterlily species *V. cruziana*, *E. ferox*, and *B. schreberi* allowed us to compare gene duplication and loss events between waterlilies and other angiosperms. Explorations of gene duplication and loss events in early angiosperms may enhance our understanding of broader gene evolution within angiosperms and provide insights into the phenotypic diversities among flowering plants. Previous studies have identified reduced numbers of resistance (R) genes in aquatic plants (Liu et al., 2021). Our study showed that the resistance-related genes *PAD4* and *ADR1* were lost, whereas *NRG1* underwent gene duplication in waterlilies. These findings suggest a unique resistance mechanism in early aquatic plants that deserves further attention. Additionally, we found that many anther dehiscence- and pollen development-related genes (i.e., orthologs of *DYT1* and *ERL1*) were lost in waterlilies; this loss may be linked to late anther dehiscence specifically observed in *Victoria*. These genes

have demonstrated relationships with anther development in other angiosperms (DeYoung et al., 2006; Alvarez-Buylla et al., 2010). Furthermore, *SABATH* genes have undergone duplication in waterlilies and may be correlated with FVOC biosynthesis in *Victoria*.

Overall, we found evidence of frequent polyploidization in waterlilies, including WGD events in Nymphaeaceae that are absent from Cabombaceae. We propose that substantial gene duplication and loss events, particularly within Nymphaeales, play an important and previously underestimated role in the evolutionary history of angiosperms.

The adaptive implications of gigantism

The dimensions of leaves and flowers in the genus *Victoria* are striking. The giant leaves of *Victoria* probably evolved as an adaptation to outcompete other species by maximizing photosynthesis potential in fast-drying pools (Box et al., 2022). Nevertheless, it has been unknown how such proportions are achieved at the molecular level. Here, we show that the accumulation of large quantities of cells in early leaf development governs ultimate organ size. Enhanced energy transportation and supply, driving a highly efficient leaf vein transport system, presumably offset the energetic costs of producing giant organs.

Regarding the flowers, their gigantic proportions may be the result of directional selection favoring the congregation of large numbers of beetles seeking heat, thus driving the evolution of flowers with increased reproductive capacity (Davis et al., 2008). Unlike most angiosperms, which have four discrete floral organs (e.g., calyx, tepals, stamens, and carpels), the flowers of waterlilies—similar to those of other early angiosperms—possess transitional floral organs (Soltis et al., 2007). Accordingly, our study indicated that the fading ABC(D)E model is involved in regulating giant flower development in *Victoria*. We found that *AGL6* plays an important role in the transition from sepals to petals. We also found that dichogamy may be regulated by *VicDDE1* and *VicMYB21*. *DDE1*, involved in filament elongation and anther dehiscence (Alvarez-Buylla et al., 2010), was highly expressed on the first day of blooming, whereas the expression of *VicMYB21*—known to regulate stamen filament development (Zhang et al., 2021)—was higher on the second day.

Most waterlilies bloom during the day and produce a bouquet of aromatic alcohols and ethers to attract bees and flies (Jiang et al., 2021). In contrast, *Victoria* blooms at night and produces strong floral scents with a high content of methyl esters (up to 52.45% of total FVOCs) to attract beetles. Uniquely among angiosperms, we found a division of labor in the inner stamens of *Victoria*, which produce FVOCs for attraction and defense-related compounds for protection. Previous studies have suggested that the SABATH methyltransferase 19 is involved in methyl ester biosynthesis (Zhao et al., 2008; Zhang et al., 2020; Jiang et al., 2021). Specifically, we observed high expression of *VicSABATHa* in inner stamen tissues on the first day of blooming (40 000 FPKM). Genes encoding enzymes of the methyl ester biosynthesis pathway that act upstream of *VicSABATHa*, such as *VicmetE* and *VicSAMs*, also ranked among the top 20 most highly expressed genes in the inner stamens, further supporting their involvement in producing a rich array of FVOCs. Biochemical assays and functional verification of VicSABATHa clarified its role in catalyzing the synthesis of short-chain fatty acid methyl esters, such as methyl butyrate, methyl tiglate, and methyl hexanoate, which are notable FVOCs produced by giant waterlilies. Compared with NcSABATH1 in *N. colorata*, VicSABATHa exhibits higher catalytic activity toward short-chain fatty acids (Figure 6E). Additionally, we detected co-expression of *Vicchitinase1* with *VicSABATHa* in the inner stamens. Glycoside hydrolase family 19 chitinases, also known as class I chitinases, play crucial roles as pathogenesis-related proteins that typically act on insect exoskeletons and fungal cell walls during herbivory and pathogen attack (Renner and Specht, 2012). Notably, chitinase-encoding genes found in the carnivorous plant *Nepenthes*, which digests insect carapaces, also belong to the class I chitinase family (Renner and Specht, 2012). *Vicchitinase1* was identified as a member of the class I chitinase gene family, suggesting a unique function in the protection of reproductive organs from damage by pollinating beetles.

METHODS

Sample collection, library establishment, and sequencing

Sample preparation for genomic sequencing and Hi-C libraries is detailed in Supplemental Table 2. Genomic sequencing of *V. cruziana* and *E. ferox* was performed using Oxford Nanopore Technologies, and DNA extraction followed our previous protocol (Zhang et al., 2020). For *N. mexicana* and *B. schreberi*, genomic sequencing was carried out via single-tube long fragment read (stLFR) technology. DNA sample preparation, DNA extraction, stLFR library construction, and sequencing were performed in accordance with published methods (He et al., 2022).

Hi-C libraries were constructed using young leaves. Leaves were fixed with formaldehyde and lysed, then, cross-linked DNA was digested overnight with MboI to obtain ~500–700 bp DNA segments. Plant organ tissues of *V. cruziana* were collected by manual dissection, immediately frozen in liquid nitrogen, and stored at −80°C before processing. Total RNA was extracted using the TRIzol reagent (Invitrogen) and sequenced to generate 100-bp-long paired-end reads on an Illumina HiSeq2000 platform (Novogene).

Genome assembly and assessment

We assembled the *V. cruziana* and *E. ferox* genomes via Nanopore long reads using wtdbg2 (Ruan and Li, 2020) with optimized parameters. The clean data were corrected using Canu v.1.5 (Koren et al., 2017), followed by three rounds of correction with Racon v.1.3.3 (Vaser et al., 2017) using long reads generated by Oxford Nanopore Technologies. Next, three additional rounds of correction were performed using Illumina short reads with Pilon v.1.2.3 (Walker et al., 2014). Finally, draft assembly genomes of *V. cruziana* and *E. ferox* were obtained.

The genomes of *N. mexicana* and *B. schreberi* were assembled from stLFR libraries using Supernova v.2.0.1. First, stLFR barcodes were transformed into 10× Genomics-compatible format using scripts available on GitHub (https://github.com/BGI-Qingdao/stlfr2supernova_pipeline). The transformed barcodes of the stLFR clean reads were then used for *de novo* assembly (Weisenfeld et al., 2017). We used GapCloser v.1.1.12 with paired-end reads to fill gaps in the assemblies (Luo et al., 2012).

Hi-C sequencing reads were mapped onto the draft assembly genomes. Based on Hi-C links and Lachesis software (https://github.com/shendurelab/LACHESIS) (Burton et al., 2013), the genomes of *V. cruziana*, *E. ferox*, and *N. mexicana* were grouped into 14, 29, and 27 chromosomes, respectively. We used RaGOO (v.1.1) (Alonge et al.,

2019) with default parameters to anchor the contigs of *B. schreberi* to 36 pseudochromosomes; the recently published *B. schreberi* genome (Lu et al., 2023) served as a reference. Each genome's assembly quality was evaluated using the BUSCO dataset embryophyta_odb10 (BUSCO v.3.0.2) (Simao et al., 2015).

Prediction and functional annotation of protein-coding genes

Protein-coding genes were predicted through *de novo*, homology-based, and evidence-based approaches. *De novo* gene prediction was performed using Augustus v.3.4 (Stanke et al., 2008), Glimmer HMM v.3.0.4 (Majoros et al., 2004), and SNAP (Korf, 2004). Homology-based prediction was conducted using GeneWise v.2.2.0, with a minimum gene model coverage of 50% (Birney et al., 2004). For evidence-based gene prediction, all RNA sequencing data were aligned to the assembled genome using Hisat 2 v.2.0.4 (Kim et al., 2015) and StringTie v.1.2.2 (Pertea et al., 2015). The cDNA sequences were further assembled using TRINITY v.2.3.2 (Grabherr et al., 2011). Protein-coding genes were functionally annotated against the public protein databases NCBI NR (http://www.ncbi.nlm.nih.gov, 20201015) and Pfam (http://pfam-legacy.xfam.org/).

Repeat analysis

To comprehensively identify repeat sequences in four Nymphaeales species, we built a *de novo* repeat library using RepeatModeler (http://repeatmasker.org/RepeatModeler.html) (Bao et al., 2015). This library was merged with a known repetitive element database (Repbase; http://www.girinst.org/repbase) to form a combined repeat database. We then used RepeatMasker with default parameters to identify repetitive sequences in the four Nymphaeales genomes (Tarailo-Graovac and Chen, 2009). We also identified full-length LTR retrotransposons and calculated their insertion times using LTR_retriever, based on the formula $T = K/2r$ (Ou and Jiang, 2018), where K represents genetic distance and r represents the mutation rate of repeat sequences.

Polyploidization analysis of waterlily genomes

To investigate the occurrence of polyploidization events in waterlily genomes, we carried out comparative genomic analyses of six waterlily species: *V. cruziana*, *N. colorata*, *N. mexicana*, *N. thermarum*, *E. ferox*, and *B. schreberi*. Homologous genes within and between genomes were identified using BLASTp with an E-value cutoff of $1e^{-5}$ (Camacho et al., 2009). We performed dot plot analyses of homologous genes within and between genomes using the whole-genome duplication integrated analysis tool (official name: WGDI; -d and -kd options). Collinear genes generated by different events were classified based on the complementarity of collinear regions in the dot plots, as well as the range of collinear gene *Ks* values. We then used a Gaussian function to fit the probability density curves of median *Ks* values within collinear blocks. Correction of evolutionary rates was conducted using a previously reported method (Wang et al., 2019; Yang et al., 2020; Bao et al., 2023).

We determined WGD events based on the combined results of homologous dot plots, peak maps, and gene trees of related species, both within and between species. The presence of clear *Ks* peak shapes helps to confirm polyploidization events; subsequent evolutionary rate calculations allow inference of the time intervals for these events. A shared WGD event was identified among the study species *V. cruziana*, *N. mexicana*, *N. colorata*, and *E. ferox*; thus, we applied a *Ks* correction method for shared events. First, if a polyploidization event is shared between species A and B, then the timing of the event should be the same across both species, and their *Ks* peak values should be equal (i.e., *Ks* A = *Ks* B). However, given that different species exhibit distinct evolutionary rates, *Ks* A and *Ks* B are not equal in practice. Assuming that after the polyploidization event, species A and B evolved at their respective rates, VA and VB, and that the ancestral lineage leading to both species evolved at a rate V up to their divergence point O, correction factors must be applied. The correction coefficient for species A is $\lambda A = V/VA$; for species B, the coefficient is $\lambda B = V/VB$. Therefore, the *Ks* AB value, corrected for differentiation between species A and B, is calculated as *Ks* AB-correction = *Ks* AB $*\lambda A*\lambda B$.

Phylogenetic tree construction and gene family evolution

We constructed a phylogenetic tree of 29 angiosperms using three gymnosperm species as outgroups. Low-copy orthologous proteins from the 29 angiosperms and the gymnosperms were identified using OrthoFinder v.2.4.0 (Emms and Kelly, 2015) and aligned using MAFFT v.7.429 with the merge method (Katoh and Standley, 2013). FastTree v.2.1 was used to generate the species tree (Price et al., 2009). We divided the 29 angiosperm species into four lineages: eudicots, monocots, magnoliids, and Nymphaeales. Statistical analyses of gene duplication and loss events that occurred within these four lineages were performed using TBtools, based on the results from OrthoFinder v.2.4.0.

Transcriptome assembly, gene expression analysis, and expression profiling using WGCNA

After the removal of adapters and low-quality reads, clean reads were mapped to the *V. cruziana* genome using Hisat2 v.2.0.4 (Kim et al., 2015) with default parameters. BAM files of uniquely mapped reads were used as input, and FPKM values were calculated to measure gene expression levels using StringTie v.2.1.4 (Kovaka et al., 2019). To identify candidate genes involved in flower development and floral scent biosynthesis, we selected several differentially expressed gene (DEG) sets for co-expression analysis. WGCNA was performed using R packages with a thresholding power of 7 to establish a gene adjacency matrix (Langfelder and Horvath, 2008). DEGs were hierarchically clustered and grouped into distinct color modules using Dynamic Hybrid Tree Cut (Langfelder and Horvath, 2008).

Identification of candidate genes and construction of phylogenetic trees

Flower development-related gene families, such as MADS-box genes, were identified using BLASTp v.2.9.0 with an E-value cutoff of 10^{-5}, utilizing homologous genes from *Arabidopsis* as query sequences (Eddy, 1998) and HMMER with the "-cut_tc" parameter (Camacho et al., 2009). Multiple sequence alignment was performed using MAFFT v.7.453 (Katoh and Standley, 2013), and phylogenetic trees were constructed using FastTree v.2.1 (Price et al., 2009).

Volatile compound measurement

To investigate floral volatile compounds in *V. cruziana*, we collected flowers on the first and second days of blooming for gas chromatography-mass spectrometry (GC-MS) analysis, following a previous study (Bao et al., 2023). Volatile compounds from *V. cruziana* flowers and transgenic *N. benthamiana* leaves were collected using headspace solid-phase microextraction. The extracted volatile compounds were thermally desorbed and transferred into an Agilent 5975-6890N GC-MS system for analysis. Compounds were identified by comparing their mass spectra with entries in the National Institute of Standards and Technology (NIST) 2008 mass spectral library and with authentic standards.

Biochemistry of VicSABATHa

Using cDNA obtained from *V. cruziana* and *N. colorata* as templates, specific primers designed based on the gene sequences of *VicSABATHa* and *NcSABATH1* were used for PCR amplification to clone the coding DNA sequences (CDSs) of *VicSABATHa* and *NcSABATH1*. Subsequently, the CDS fragments were individually cloned into the pESI-Blunt vector to obtain target fragments for the construction of expression vectors. These target fragments were then ligated into the pEAQ-HT vector (previously digested with NruI and XhoI) to construct eukaryotic expression vectors (HT-VicSABATH1 and HT-NcSABATH1) for further experiments. For prokaryotic expression, the CDS fragments were ligated into the pET32a

vector after digestion with BamHI and EcoRI to generate the pET32a-VicSABATH1 and pET32a-NcSABATH1 constructs.

For *in vitro* enzyme activity assays, the prokaryotic expression vectors (pET32a-VicSABATHa and pET32a-NcSABATH1) were transformed into *Escherichia coli* BL21-competent cells to produce soluble recombinant proteins. These proteins were subjected to *in vitro* enzyme activity assays using a specific reaction system. Reaction products were adsorbed onto polydimethylsiloxane solid-phase microextraction fibers and sealed at 30°C for 2 h; volatile compounds were then desorbed and detected using GC-MS to analyze the enzyme reaction products.

To explore the function of VicSABATHa and NcSABATH1, we carried out transient transfection experiments in *Nicotiana benthamiana*. The pEAQ-HT empty vector and the eukaryotic expression vectors (HT-VicSABATH1 and HT-NcSABATH1) were transformed into *Agrobacterium tumefaciens* GV3101 to prepare infiltration solutions. These solutions were injected into the abaxial side of tobacco leaf epidermis. Suitable leaves were selected for infiltration, marked, and harvested 4 days post-infiltration. The infiltrated leaf samples were weighed, cut, and placed in sealed glass bottles for 24 h at 22°C. Volatile compounds released from the tobacco leaves were analyzed by GC-MS.

For enzyme kinetics experiments, the optimal reaction conditions—including substrate concentration, temperature, pH, and reaction time—were determined. Enzyme-catalyzed reactions were performed using at least five gradients of substrate concentrations and at least five linear time points. Standard curves were plotted, with the slope representing the initial reaction velocity [V] at a specific substrate concentration [S]. Reciprocal values of [S] and [V] were plotted as 1/[S] versus 1/[V], yielding the slope (k) and intercept (b). These values were then used to calculate enzyme kinetics parameters via the following formulas: [E] = protein mass (μg) / (reaction volume [μl] × protein size [kDa]); $K_{M\ =}$ k/b; V_{max} = 1/b; kcat = V_{max}/[E].

ACCESSION NUMBERS

The data supporting this work are available within the paper and its supplemental information. Raw reads for *N. Mexicana* and *B. schreberi* are available at China National GeneBank DataBase (https://db.cngb.org/) under project number CNP0001360. The raw reads for *V. cruziana and E. ferox* are available at the BIG Data Center (https://ngdc.cncb.ac.cn/) under project number PRJCA040089. Genome assemblies and annotations can be accessed via Figshare (https://doi.org/10.6084/m9.figshare.22099592.v2). Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

FUNDING

This work was supported by the National Natural Science Foundation of China (32272750 and 31970246), the Scientific and Technological Talents and Platform Program of Yunnan Province (Academician Expert Workstation, 202305AF150165), and the ''Pioneer'' and ''Leading Goose'' R&D Program of Zhejiang (2023C02028).

ACKNOWLEDGMENTS

We are very grateful to Douglas E. Soltis and Pamela S. Soltis for their revisions and suggestions. No conflict of interest is declared.

AUTHOR CONTRIBUTIONS

L.Z. conceived and designed the project. X.W., Yu L., H.S., X.C., X.S., S. S., Y.F., D.C., F.C., Yue L., Qian G., Qiang G., and Q.W. performed the experiments and analyzed the results. X.W., Yu L., and L.Z. wrote the manuscript. L.Z., Yong L., Z.W., H.K., H.L., X.G., X.Z., and T.C. revised the manuscript. All authors read and approved the final version of the manuscript.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

Received: September 25, 2024
Revised: March 17, 2025
Accepted: April 14, 2025
Published: April 16, 2025

REFERENCES

Albalat, R., and Cañestro, C. (2016). Evolution by gene loss. Nat. Rev. Genet. **17**:379–391.

Alemán-Báez, J., Acevedo-Zamora, J.F., Bucher, J., Cai, C., Voorrips, R.E., and Bonnema, G. (2024). Expression changes of miRNA-regulated genes associated with the formation of the leafy head in cabbage. Hortic. Plant J. **10**:1007–1019.

Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F.J., Lippman, Z.B., and Schatz, M.C. (2019). RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. **20**:224.

Alvarez-Buylla, E.R., Benítez, M., Corvera-Poiré, A., Chaos Cador, A., de Folter, S., Gamboa de Buen, A., Garay-Arroyo, A., García-Ponce, B., Jaimes-Miranda, F., Pérez-Ruiz, R.V., et al. (2010). Flower development. Arabidopsis Book **8**:e0127.

Bao, T., Kimani, S., Li, Y., Li, H., Yang, S., Zhang, J., Wang, Q., Wang, Z., Ning, G., Wang, L., and Gao, X. (2023). Allelic variation of terpene synthases drives terpene diversity in the wild species of the Freesia genus. Plant Physiol. **192**:2419–2435.

Bao, W., Kojima, K.K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob. DNA **6**:11.

Becker, A., and Günter, T. (2003). The major clades of mads-box genes and their role in the development and evolution of flowering plants. Mol. Phylogenet. Evol. **29**:464–489.

Bernhardt, P. (2000). Convergent evolution and adaptive radiation of beetle-pollinated angiosperms. Plant Syst Evol **222**:293–320.

Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and Genomewise. Genome Res. **14**:988–995.

Bowers, J.E., Chapman, B.A., Rong, J., and Paterson, A.H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature **422**:433–438.

Box, F., Erlich, A., Guan, J.H., and Thorogood, C. (2022). Gigantic floating leaves occupy a large surface area at an economical material cost. Sci. Adv. **8**:eabg3790.

Brebbia, C.A., and Assoc, E.S.T. (2007). Design and Nature: Comparing design in nature with science and engineering. Ind. Ceram. **27**.

Buggs, R.J.A. (2017). The deepening of Darwin's abominable mystery. Nat. Ecol. Evol. **1**:169.

Burton, J.N., Adey, A., Patwardhan, R.P., Qiu, R., Kitzman, J.O., and Shendure, J. (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat. Biotechnol. **31**:1119–1125.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinf. **10**:421.

Chen, F., Liu, X., Yu, C., Chen, Y., Tang, H., and Zhang, L. (2017a). Water lilies as emerging models for Darwin's abominable mystery. Hortic. Res. **4**:17051.

Chen, F., Zhang, X., Liu, X., and Zhang, L. (2017b). Evolutionary analysis of MIKCc-type Mads-box genes in gymnosperms and angiosperms. Front. Plant Sci. **8**:895.

Davis, C.C., Endress, P.K., and Baum, D.A. (2008). The evolution of floral gigantism. Curr. Opin. Plant Biol. **11**:49–57.

DeYoung, B.J., Bickle, K.L., Schrage, K.J., Muskett, P., Patel, K., and Clark, S.E. (2006). The CLAVATA1-related BAM1, BAM2 and BAM3 receptor kinase-like proteins are required for meristem function in Arabidopsis. Plant J. **45**:1–16.

Doyle, J.A. (2012). Molecular and fossil evidence on the origin of angiosperms. Annu. Rev. Earth Planet Sci. **40**:301–326.

Dudareva, N., Klempien, A., Muhlemann, J.K., and Kaplan, I. (2013). Biosynthesis, function and metabolic engineering of plant volatile organic compounds. New Phytol. **198**:16–32.

Eddy, S.R. (1998). Profile hidden Markov models. Bioinformatics **14**:755–763.

Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. **16**:157.

Endress, P.K. (2010). The evolution of floral biology in basal angiosperms. Philos. Trans. R. Soc. Lond. B Biol. Sci. **365**:411–421.

Ervik, F., and Knudsen, J.T. (2003). Water lilies and scarabs: faithful partners for 100 million years? Biol. J. Linn. Soc. **80**:539–543.

Friis, E.M., Crane, P.R., and Pedersen, K.R. (2018). Tanispermum, a new genus of hemi-orthotropous to hemi-anatropous angiosperm seeds from the Early Cretaceous of eastern North America. Am. J. Bot. **105**:1369–1388.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. **29**:644–652.

Grover, A. (2012). Plant chitinases: genetic diversity and physiological roles. Crit. Rev. Plant Sci. **31**:57–73.

He, Z., Feng, X., Chen, Q., Li, L., Li, S., Han, K., Guo, Z., Wang, J., Liu, M., Shi, C., et al. (2022). Evolution of coastal forests based on a full set of mangrove genomes. Nat. Ecol. Evol. **6**:738–749.

Jana, B.R. (2018). Flower characteristics and pollination behavior of *Euryale ferox* (Salisb.). Am. J. Plant Sci. **9**:722–731.

Jiang, Y., Liu, G., Zhang, W., Zhang, C., Chen, X., Chen, Y., Yu, C., Yu, D., Fu, J., and Chen, F. (2021). Biosynthesis and emission of methyl hexanoate, the major constituent of floral scent of a night-blooming water lily *Victoria cruziana*. Phytochemistry **191**:112899.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. Nature **473**:97–100.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. **30**:772–780.

Keinan, O., Valentine, J.M., Xiao, H., Mahata, S.K., Reilly, S.M., Abu-Odeh, M., Deluca, J.H., Dadpey, B., Cho, L., Pan, A., et al. (2021). Glycogen metabolism links glucose homeostasis to thermogenesis in adipocytes. Nature **599**:296–301.

Kim, D., Landmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods **12**:357–U121.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. **27**:722–736.

Korf, I. (2004). Gene finding in novel genomes. BMC Bioinf. **5**:59.

Kovaka, S., Zimin, A.V., Pertea, G.M., Razaghi, R., Salzberg, S.L., and Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. **20**:278.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. **9**:559.

Liu, Y., Zeng, Z., Zhang, Y.M., Li, Q., Jiang, X.M., Jiang, Z., Tang, J.H., Chen, D., Wang, Q., Chen, J.Q., and Shao, Z.Q. (2021). An angiosperm NLR Atlas reveals that NLR gene reduction is associated with ecological specialization and signal transduction component deletion. Mol. Plant **14**:2015–2031.

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

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience **1**:18.

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

Ng, M., and Yanofsky, M.F. (2001). Function and evolution of the plant MADS-box gene family. Nat. Rev. Genet. **2**:186–195.

Ou, S., and Jiang, N. (2018). LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. Plant Physiol. **176**:1410–1422.

Paez, S., Kraus, R.H.S., Shapiro, B., Gilbert, M.T.P., Jarvis, E.D., Al-Ajli, F.O., Ceballos, G., Crawford, A.J., Fedrigo, O., et al.; Vertebrate Genomes Project Conservation Group (2022). Reference genomes for conservation. Science **377**:364–366.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. **33**:290–295.

Povilus, R.A., DaCosta, J.M., Grassa, C., Satyaki, P.R.V., Moeglein, M., Jaenisch, J., Xi, Z., Mathews, S., Gehring, M., Davis, C.C., and Friedman, W.E. (2020). Water lily (*Nymphaea thermarum*) genome reveals variable genomic signatures of ancient vascular cambium losses. Proc. Natl. Acad. Sci. USA **117**:8649–8656.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol. Biol. Evol. **26**:1641–1650.

Renner, T., and Specht, C.D. (2012). Molecular and functional evolution of class I chitinases for plant carnivory in the caryophyllales. Mol. Biol. Evol. **29**:2971–2985.

Rodriguez, R.E., Mecchia, M.A., Debernardi, J.M., Schommer, C., Weigel, D., and Palatnik, J.F. (2010). Control of cell proliferation in Arabidopsis thaliana by microRNA miR396. Development **137**:103–112.

Ruan, J., and Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. Nat. Methods **17**:155–158.

Sargent, R.D., and Otto, S.P. (2004). A phylogenetic analysis of pollination mode and the evolution of dichogamy in angiosperms. Evol. Ecol. Res. **6**:1183–1199.

Schimpf, N.G., Terblanche, J.S., Smit, M.F., and Matthews, P.G.D. (2017). Respiration, thermogenesis, and thermoregulation of *Victoria cruziana* flowers. Aquat. Bot. **138**:37–44.

Seymour, R.S., and Matthews, P.G.D. (2006). The role of thermogenesis in the pollination biology of the Amazon waterlily *Victoria amazonica*. Ann. Bot. **98**:1129–1135.

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

Smith, L.T., Magdalena, C., Przelomska, N.A.S., Perez-Escobar, O.A., Melgar-Gomez, D.G., Beck, S., Negrao, R., Mian, S., Leitch, I.J.,

Dodsworth, S., et al. (2022). Revised species delimitation in the giant water lily Genus (Nymphaeaceae) Confirms a New Species and Has Implications for Its Conservation. Front. Plant Sci. **13**:883151.

Soltis, D.E., Chanderbali, A.S., Kim, S., Buzgo, M., and Soltis, P.S. (2007). The ABC model and its applicability to basal angiosperms. Ann. Bot. **100**:155–163.

Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics **24**:637–644.

Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics Chapter **4**:Unit4.10.

Thien, L.B., Bernhardt, P., Devall, M.S., Chen, Z.D., Luo, Y.B., Fan, J.H., Yuan, L.C., and Williams, J.H. (2009). Pollination biology of basal angiosperms (Anita Grade). Am. J. Bot. **96**:166–182.

Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. **27**:737–746.

Wagner, A.M., Krab, K., Wagner, M.J., and Moore, A.L. (2008). Regulation of thermogenesis in flowering Araceae: the role of the alternative oxidase. Biochim. Biophys. Acta **1777**:993–1000.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One **9**:e112963.

Wang, J., Yuan, J., Yu, J., Meng, F., Sun, P., Li, Y., Yang, N., Wang, Z., Pan, Y., Ge, W., et al. (2019). Recursive paleohexaploidization shaped the durian genome. Plant Physiol. **179**:209–219.

Weisenfeld, N.I., Kumar, V., Shah, P., Church, D.M., and Jaffe, D.B. (2017). Direct determination of diploid genome sequences. Genome Res. **27**:757–767.

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

Yang, X., Zhao, X., Dai, Z., Ma, F., Miao, X., and Shi, Z. (2021). OsmiR396/growth regulating factor modulate rice grain size through direct regulation of embryo-specific miR408. Plant Physiol. **186**:519–533.

Yang, Y., Sun, P., Lv, L., Wang, D., Ru, D., Li, Y., Ma, T., Zhang, L., Shen, X., Meng, F., et al. (2020). Prickly waterlily and rigid hornwort genomes shed light on early angiosperm evolution. Nat. Plants **6**:215–222.

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

Zhang, X., He, Y., Li, L., Liu, H., and Hong, G. (2021). Involvement of the R2R3-MYB transcription factor MYB21 and its homologs in regulating flavonol accumulation in Arabidopsis stamen. J. Exp. Bot. **72**:4319–4332.

Zhao, N., Ferrer, J.L., Ross, J., Guan, J., Yang, Y., Pichersky, E., Noel, J.P., and Chen, F. (2008). Structural, biochemical, and phylogenetic analyses suggest that indole-3-Acetic acid methyltransferase is an evolutionary ancient member of the SABATH family. Plant Physiol. **146**:323–324.