



OPEN ACCESS

EDITED BY

Adrian Christopher Brennan,
Durham University,
United Kingdom

REVIEWED BY

Tangren Cheng,
Beijing Forestry University,
China
Shahid Ullah Khan,
Huazhong Agricultural University,
China
Chaoying He,
Institute of Botany (CAS), China

*CORRESPONDENCE

Xingyu Yang
sanmu109@163.com
Tao Shi
shitao323@wbgcas.cn

SPECIALTY SECTION

This article was submitted to
Plant Systematics and Evolution,
a section of the journal
Frontiers in Plant Science

RECEIVED 20 June 2022

ACCEPTED 23 August 2022

PUBLISHED 23 September 2022

CITATION

Gao Z, Liang Y, Wang Y, Xiao Y, Chen J,
Yang X and Shi T (2022) Genome-wide
association study of traits in sacred lotus
uncovers MITE-associated variants
underlying stamen petaloid and petal
number variations.
Front. Plant Sci. 13:973347.
doi: 10.3389/fpls.2022.973347

COPYRIGHT

© 2022 Gao, Liang, Wang, Xiao, Chen,
Yang and Shi. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Genome-wide association study of traits in sacred lotus uncovers MITE-associated variants underlying stamen petaloid and petal number variations

Zhiyan Gao^{1,2,3}, Yuting Liang⁴, Yuhan Wang⁵, Yingjie Xiao⁶,
Jinming Chen^{1,2}, Xingyu Yang^{4*} and Tao Shi^{1,2*}

¹Key Laboratory of Aquatic Botany and Watershed Ecology, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, China, ²Center of Conservation Biology, Core Botanical Gardens, Chinese Academy of Sciences, Wuhan, China, ³College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China, ⁴Wuhan Institute of Landscape Architecture, Wuhan, China, ⁵Wuhan Institute of Design and Sciences, Wuhan, China, ⁶National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, China

Understanding the genetic variants responsible for floral trait diversity is important for the molecular breeding of ornamental flowers. Widely used in water gardening for thousands of years, the sacred lotus exhibits a wide range of diversity in floral organs. Nevertheless, the genetic variations underlying various morphological characteristics in lotus remain largely unclear. Here, we performed a genome-wide association study of sacred lotus for 12 well-recorded ornamental traits. Given a moderate linkage disequilibrium level of 32.9kb, we successfully identified 149 candidate genes responsible for seven flower traits and plant size variations, including many pleiotropic genes affecting multiple floral-organ-related traits, such as *NnKUP2*. Notably, we found a 2.75-kb presence-and-absence genomic fragment significantly associated with stamen petaloid and petal number variations, which was further confirmed by re-examining another independent population dataset with petal number records. Intriguingly, this fragment carries MITE transposons bound by siRNAs and is related to the expression differentiation of a nearby candidate gene between few-petalled and double-petalled lotuses. Overall, these genetic variations and candidate genes responsible for diverse lotus traits revealed by our GWAS highlight the role of transposon variations, particularly MITEs, in shaping floral trait diversity.

KEYWORDS

GWAS, MITEs, sacred lotus, stamen petaloid, petal number

Introduction

Understanding the genetic and molecular basis of phenotypic diversity among individuals within species is a long-standing problem in evolutionary developmental biology and population genetics (Huang and Han, 2014; Payne and Wagner, 2019). Previously, it was widely accepted that genetic variation, including single nucleotide

variants and structural variants, in protein-coding regions directly alters protein structures and functions to control developmental complexity and phenotypic variation (Syvanen, 2001; Shastry, 2009). Currently, growing empirical evidence highlights the predominant role of variations in non-coding regulatory regions in affecting morphological variations in animals and plants (Glazier et al., 2002; Carroll, 2005; Hoekstra and Coyne, 2007; Romero et al., 2012; Signor and Nuzhdin, 2018). To detect significant associations between genotypes and phenotypic variance in diverse populations, a genome-wide association study (GWAS) based on linkage disequilibrium (LD) has become a powerful approach to detect many natural allelic variations simultaneously in a single study (Hayes, 2013). GWASs of different crops have discovered critical molecular markers containing single nucleotide polymorphism (SNP) loci and candidate gene regions associated with structural and physiological traits in maize (Tian et al., 2011), tomato (Lin et al., 2014), soybean (Zhang et al., 2015), and rice (Yano et al., 2016). Intriguingly, many recent GWASs of different plant systems have also found that transposon insertions, particularly miniature inverted-repeat transposable elements (MITEs), play an important role in trait diversity by affecting the expression or structural change of nearby genes. In rice, a transposon-insertion-polymorphisms-GWAS revealed that the insertions of four significant MITEs located inside key genes or in their close vicinity were associated with six grain-related phenotypes, including grain width (Castanera et al., 2021). In another GWAS example, a 125 bp MITE insertion in the promoter region of the *TaVSRI-B* gene suppressed the expression of this gene, leading to longer and thinner roots of wheat at the booting stage (Wang et al., 2021). In addition, a GWAS based on presence and absence variations detected a 621 bp MITE that was inserted in the promoter region and enhanced the expression of *FLC*, thus enhancing the vernalization of winter-type oilseed rapes (Song et al., 2020). These pieces of evidence seem to show that transposon variations are one of the driving forces underlying phenotypic diversity in plant adaptation and domestication.

Floral traits, such as flower shapes in *Arabidopsis* (Takeda et al., 2013), flower sizes in *Penstemon* (Parachnowitsch and Kessler, 2010) and flower colors in *Petunia* (Tornielli et al., 2009), are diverse and often desirable for horticulture. Understanding the genetic variations underlying floral morphological differentiation is a crucial task for both plant developmental biology and horticulture research, particularly for ornamental flowers, such as chrysanthemum, rose and sacred lotus. Although parts of the biosynthesis pathways regulating floral development have been verified to be relatively conserved across model plants (Litt and Kramer, 2010; Ruiz-Sola and Rodriguez-Concepcion, 2012; Saito et al., 2013), several studies on non-model plant species are needed to verify and investigate the multitude of functional alleles related to flowering time, floral shapes, pigmentation and petal numbers based on GWAS. For example, an SNP in the *PPR* gene was linked to flowering time leading to early- and late-flowering pearl millets (Diack et al., 2020). In another example, both known and novel genes associated with petal size were identified in rapeseed (Qian

et al., 2021). Some flower color variants were also identified by GWAS case studies. In *Prunus mume* and rose, several known candidate genes (*4CL*) were verified, and some new candidate genes (*CDC6*) were found to affect anthocyanin and carotenoid concentrations (Schulz et al., 2016; Zhang et al., 2018). In addition, modeling multiple flower traits with polymorphisms together in one study could improve the prediction accuracy of key genes underlying trait variations when compared with the study of one trait independently, especially for correlated plant development traits. In *Petunia*, there were significant correlations between flower length, flower diameter, flower number and florescence, and multiple SNP loci were colocalized in four intervals on four linkage groups controlling various traits (Cao et al., 2018). Therefore, the study of multiple flower traits in a single GWAS could help to comprehensively reveal the mechanisms underlying the diversity of complex developmental traits.

Lotus (*Nelumbo nucifera* Gaertn or sacred lotus) is a famous and traditional aquatic flower with significant cultural and religious symbolization in ancient civilizations, such as China and India. It exhibits diverse and attractive floral traits with thousands of years of gardening (Wang, 2005). Lotus can be classified into three main groups according to their primary utilizations, namely rhizome lotus, seed lotus, and flower lotus (Wang, 2005; Lin et al., 2019). Flower lotus has been widely cultivated in Asia because of its diverse flower coloration and complex floral morphologies (Guo, 2008). The domestication history of flower lotus from few-petalled red flowers to double-petalled, duplicate-petalled, and all-double-petalled flowers can be traced back to China's Jin dynasty. With the deepening of cultural communication and the development of cultivation techniques in lotus, the all-double-petalled lotus was cultivated in the Northern and Southern Dynasties of China (Wang, 2005; Guo, 2008; Liu et al., 2020). A broad spectrum of flower colors and patterns from white to red exists in lotus, which is associated with the amount and types of anthocyanin affected by various epigenetic regulators and the expression of structural genes (such as *CHI*, *UFGT*, *ANS*, *FLS*, *OMTs*, and *GST*; Deng et al., 2015; Sun et al., 2016; Wang et al., 2016; Zhu et al., 2019) and transcription factors (such as *MYB5* and *TT8*; Sun et al., 2016; Deng et al., 2021). Generally, the types of lotus flowers vary from few-petalled flowers to various double-petalled flowers because flowers with stamen petaloid and carpel petaloid fail to form normal floral patterns. These processes have been predicted to be influenced by several hormone-related candidate genes and transcription factors (TFs; Lin et al., 2018), and epigenetic regulation (methylation and miR172 family; Lin et al., 2019; Zhang Y. et al., 2020). The chromosomal-level genome assembly of lotus cv. 'China Antique' (Ming et al., 2013; Shi et al., 2020) provides opportunities for the population, expression and genomic studies of its natural variations, such as in rhizome enlargement (Yang et al., 2015; Gao et al., 2021; Huang et al., 2021; Li et al., 2021a), seed development (Li et al., 2018) and flower development (Lin et al., 2019). Using a reference genome and transcriptomic analysis, some lotus genes were also predicted to control the flowering time of lotus by participating in the photoperiod, vernalization and gibberellic acid pathways (Yang et al.,

2014; Zhang et al., 2021). Although previous transcriptomic or methylation genomic studies revealed the molecular mechanisms underlying a single trait for flower development, no systematic study has determined the genetic architecture of multiple correlated traits in a population using GWAS in lotus. In this study, to better uncover the genetic variants and molecular mechanisms underlying the various traits in the cultivated flower lotus, we comprehensively performed GWAS with the generalized linear model (GLM). After further investigation of the variants in gene structure, the expression patterns of candidate genes, and the transposable elements (TEs) and small RNAs associated with floral diversity and plant size, we provide insights into the genetic basis underlying crucial ornamental trait variations of this traditional flower.

Materials and methods

Plant materials

A total of 88 cultivated lotus accessions with abundant diversity in phenotypic traits were grown in pots (50 cm × 50 cm) under the same conditions at the Wuhan Institute of Landscape Architecture (114°52' E, 30°50' N), Wuhan, China. Twelve crucial phenotypic traits, including flowering time, population florescence, flower density, petal density, flower type, stamen petaloid, flower color, flower shape, flower diameter, maximum petal length, maximum petal width, and plant size, were investigated and confirmed by at least 3 years of observations during long-term cultivation, and used for GWAS (Supplementary Table S1; Wang, 2005; Wang and Zhang, 2005; Guo, 2008; Zhang and Wang, 2011). The rationale of choosing these 88 lotus cultivars in our GWAS is because (1) they are registered cultivars while registration requires at least three consecutive years of observation on stable traits, (2) they have been cultivated and validated for trait stability for at least 3 years and only cultivars with traits consistent with registration records were remained for conservation, and (3) their key phenotypes we focused on this study have been already recorded in 'Lotus Flower Cultivars In China' (Wang, 2005), 'Colored Illustration of Lotus Cultivars in China' (Wang and Zhang, 2005), 'Cultivation Of Lotus (*Nelumbo Nucifera* Gaertn. Ssp. *Nucifera*) And Its Utilization In China' (Guo, 2008), and 'New Lotus Flower Cultivars in China I' (Zhang and Wang, 2011). All trait data in GWAS were based on the records of the books and our validations, and deposited in *Nelumbo Genome Database* (Shi et al., 2020; Li et al., 2021b).¹

SNP calling, validation, and annotation

To explore genetic variations in this lotus germplasm, the resequencing data for the abovementioned 88 accessions were

previously sequenced in an Illumina platform with more than 10× coverage depth for each sample and deposited in *Nelumbo Genome Database* (Shi et al., 2020; Li et al., 2021b). Resequencing reads were mapped to the reference lotus genome 'China Antique' from our *Nelumbo Genome Database* (Shi et al., 2020; Li et al., 2021b) with BWA (v0.7.12) using the default parameters (Li, 2013). Nearly 99.14% of clean reads were mapped to the reference genome, with an average sequencing depth of 14.93-fold (Supplementary Table S2). These mapping outputs of the 88 accessions in SAM format were then converted into the BAM format and coordinate sorted using SAMtools (version 0.1.19)² and Picard (version 2.0.1).³ Subsequently, SNP calling was carried out *via* the Genome Analysis Toolkit (GATK 4; McKenna et al., 2010) using a standard pipeline⁴ to obtain single-sample gvcf files. High-quality SNPs in the joint genotyping data were retained for subsequent analyses by restricting the missing rate (–max-missing 1), minor allele frequency (–MAF 0.05), number of alleles (–min-alleles 2, –max-alleles 2) and read depth (–minDP 3, –maxDP 100, –min-meanDP 3) for each SNP locus. These high-quality SNPs were annotated using SnpEff⁵ based on the reference genome.

Population genetic analyses

The high-quality SNPs with an $r^2 > 0.5$ (pairwise coefficient) within 100 kb windows were pruned to avoid the effects of excessive LD between adjacent SNP loci using Plink 1.9 (Purcell et al., 2007). PopLDdecay v3.31 with MaxDist set at 1,000 was used to calculate and plot the LD decay of each group (Zhang et al., 2019). The population genetic structure was determined by ADMIXTURE with cross-validation (CV) of the hypothetical number of populations (K) ranging from 1 to 10 based on a pruned set of SNPs (Alexander et al., 2009). The population structure was also assessed by principal component analysis (PCA) with Genome-Wide Complex Trait (GCTA) using pruned SNPs (Yang et al., 2011). To further provide insight into the population's genetic structure, the phylogenetic relationships among the 88 cultivated lotus samples were analyzed using SNPhylo software (Lee et al., 2014).

Genome-wide association and candidate gene analyses

A GWAS was performed with 455,292 high-quality non-missing SNPs and 12 traits from 88 accessions in TASSEL (v5.2.44)⁶ based on a generalized linear model (GLM). The GWAS results were

¹ <http://nelumbo.biocloud.net>

² <https://github.com/samtools/samtools>

³ <https://github.com/broadinstitute/picard>

⁴ <https://www.broadinstitute.org/gatk/>

⁵ <http://snpeff.sourceforge.net/>

⁶ <https://tassel.bitbucket.io/>

visualized in the form of Manhattan and QQ plots using R (v3.31). The significantly associated SNP loci were estimated based on significance level (*, $\alpha=0.05$, $-\log_{10} p=6.96$, p value $<1.106378e-07$) and corrected by Benjamini–Hochberg correction to reduce false-positive associations. According to the LD blocks of the association population, genes within ± 50 kb of the significantly associated SNP loci were identified. Homologous genes in *Nelumbo* were annotated with the highest e-value based on the TAIR10 dataset⁷ using BlastP v.2.2.31. The functions of homologous and nonhomologous genes were found in the *Arabidopsis* and *Nelumbo* databases. The expression profiles of candidate genes were collected from the *Nelumbo Genome Database*. The phylogenetic relationships of these orthologous protein sequences among lotus, *Arabidopsis* and *Oryza sativa* were constructed by the neighbor-joining (NJ) approach with 500 bootstrap replicates using MEGA v.7.0 after alignment with MUSCLE with default settings. The gene networks were visualized using Cytoscape v3.5.1 based on a weighted gene coexpression network (WGCNA) from the *Nelumbo Genome Database* to depict the relationships between genes. In addition, known lncRNAs, small RNAs, and TEs in previous studies were also identified by Bedtools (Shi et al., 2017, 2020).

To verify the association between the stamen petaloid trait and genotype variants, the sequencing coverages and genotypes of the deletion region on chromosome 5 across 88 lotus accessions examined in our study and 136 selected lotus accessions from Liu et al. (2020; with an Illumina sequencing coverage depth > 5) were calculated by Bedtools and Delly, respectively (Quinlan and Hall, 2010; Rausch et al., 2012). In addition, transcriptome datasets from petals of double-petaloid lotus ('Fenhonglingxiao'; Lin et al., 2018) and few-petaloid lotus ('China Antique') from the *Nelumbo Genome Database* were used to find expression differences between them for candidate genes close to the deletion region with StringTie (Pertea et al., 2016) and DESeq2 ($|\log_2FC| > 1$ and $P_{adj} < 0.05$; Love et al., 2014).

Results

Genome-wide SNP detection and population parameters of flower lotus

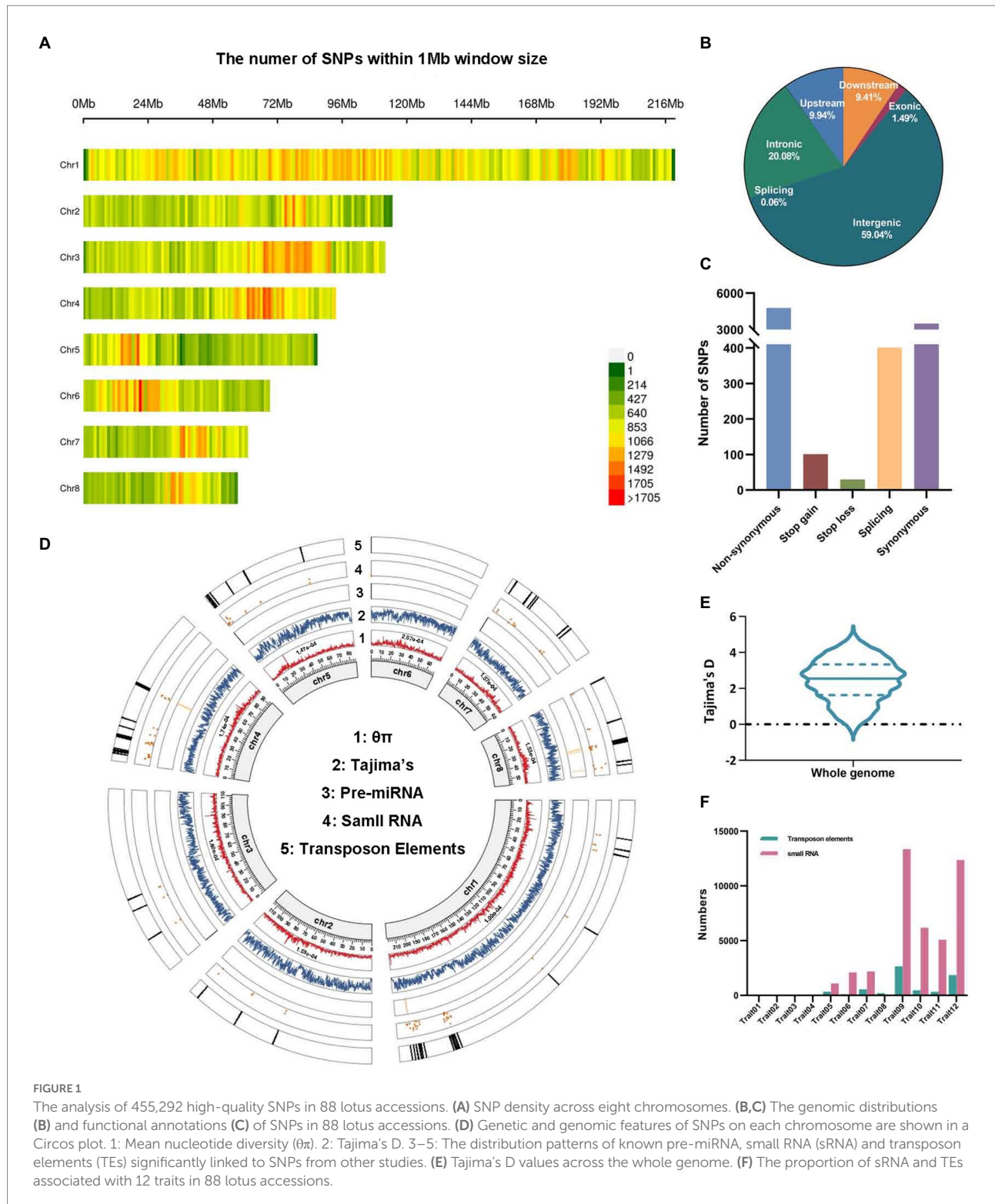
To ensure the accuracy of GWAS, it is important to select an effective population with sufficient genotypic and phenotypic diversity. A final set of 455,292 high-quality SNPs were detected based on the resequencing of 88 cultivated lotus individuals with diverse recorded traits, which were almost uniformly distributed across the eight lotus chromosomes (Figure 1A; Supplementary Tables S1, S2). Among these SNPs, 230,995 (40.96%) were located within genes, and the remainder were located in intergenic regions (59.04%; Figure 1B). In the coding regions, more than 5,300 SNPs (1.2% of the total) were predicted as

nonsynonymous (4,771), stop-gain (101), stop-loss (30), and splicing (401) variants that have a potential effect on amino acid changes, elongated transcripts or premature stop codons (Figure 1C). This percentage is lower than those for watermelon (5.6%; Guo et al., 2019) and rice (10.1%; Yano et al., 2016) but close to those for maize (1.9% Jiao et al., 2012) and cotton (1.3%; Ma et al., 2018). Furthermore, 14,435 genes were found to have SNPs in their coding sequences (CDSs). In contrast, 20,051 genes without SNP variations were found, suggesting that their CDS was highly conserved. The above results demonstrate that the SNP dataset in our study can provide abundant genetic markers for resolving the genetic mechanisms underlying various traits through GWAS.

We further assessed genome-wide nucleotide diversity ($\theta\pi$), revealing relatively higher average nucleotide diversity ($\theta\pi$) on chromosome 6 ($\theta\pi=2.07e-04$) and chromosome 1 ($\theta\pi=1.99e-04$) than on the other chromosomes across the 88 cultivated lotuses (Figure 1D). In addition, Tajima's D was greater than zero, suggesting overall balancing selection or sudden contraction of the lotus population size, which likely preserved multiple beneficial alleles (Figures 1D,E).

Population structure and PCA were performed to evaluate genetic differentiation among samples based on 155,564 SNPs after removing redundant SNPs in LD blocks. These 88 accessions were predicted to be derived from a single ancestral population, as the lowest cross-validation error was found when $K=1$ (Supplementary Figure S1A). In addition, our PCA plot showed a continuous distribution of lotus individuals without distinct clusters, which was concordant with our population structure analysis (Supplementary Figure S1B). Besides, the average proportion of heterozygosity for the whole lotus population was 0.224. 74 out of 88 lotus individuals showed observed heterozygosity ranging from 0.20 to 0.30, indicating that our population had high genetic diversity owing to extensive hybridization during breeding (Supplementary Figure S1C). The phylogenetic tree indicated that the 88 lotus accessions were indeed from one population because of extremely low bootstrap values (lower than 27) in early-branching nodes, which cannot confidently divide individuals into different populations (Supplementary Figure S2). Approximately 48.86% of the bootstrap values in this phylogenetic tree were less than 80, indicating that the phylogenetic tree had low statistical support (Supplementary Figure S1D). Therefore, the overall cultivated flower lotuses did not show strong population structure or population divergence, indicating that population structure might not affect GWAS analysis. Importantly, half of the pairwise correlation coefficient (r^2) maximum value ($r^2=0.42$) at 32.9 kb was taken as the critical value for all lotus accessions after calculating the extent of LD (Supplementary Figure S1E), which was lower than that for the flower lotus groups (58 kb) from a previous lotus study (Liu et al., 2020). This result suggested that our lotus population might have more extensive recombination through hybridization and crossing, leading to shorter linkage blocks. All the above genetic indices supported that the 88 cultivated lotuses were suitable for a GWAS because of their moderate LD decay rate, low population structure and high phenotypic diversity.

⁷ <https://www.arabidopsis.org/>



Genotypic characterization and SNP loci associated with traits

Here, we investigated the 12 key important traits of 88 lotus accessions (Figure 2; Supplementary Table S1). The seven quantitative traits, including flower time, flower density and

flower type, showed relatively normal distributions. However, the remaining quality traits, like stamen development and flower color, varied markedly among these samples (Figure 2A). The Pearson correlation analysis revealed that there were moderate to strong positive correlations between stamen petaloid and flower type ($r=0.64$, value of $p=2.682e-05$), between maximum petal

length and plant size ($r=0.65$, value of $p=2.597e-05$), between flower diameter and plant size ($r=0.71$, value of $p=8.978e-06$), between maximum petal width and flower diameter ($r=0.83$, value of $p=3.794e-09$), between maximum petal length and flower diameter ($r=0.91$, p value= $4.203e-11$), and between maximum petal width and maximum petal length ($r=0.93$, value of $p=2.422e-11$), which suggested that some traits are strongly linked during lotus flower or plant development (Figure 2B).

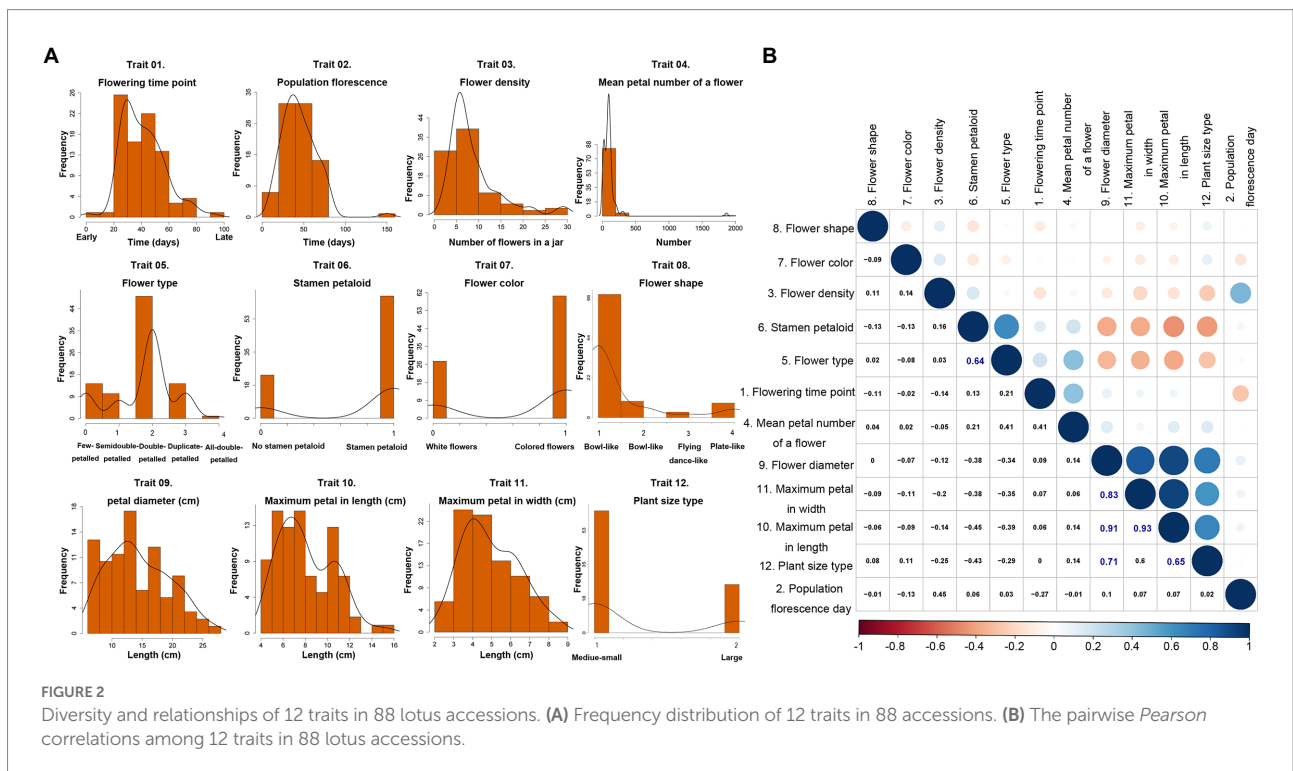
Furthermore, using 455,292 high-quality SNPs, a GWAS analysis for these 12 traits was carried out using a generalized linear model (GLM) with controlling for population structure ($K=1$). QQ-plots that displayed the distribution of calculated p value were similar to the theoretical p value in the expected diagonal, indicating the accuracy of these GWAS results (Figure 3; Supplementary Figure S3). Intriguingly, 10 out of the 12 traits had a total of 438 significantly associated SNPs given the significance threshold after excluding the repeated SNP loci based on GLM (value of $p < 1.106378e-07$; Figure 3; Supplementary Figure S3, Supplementary Table S3). Of these 438 SNPs, the number of SNPs in non-coding regions (72.15%) was two times higher than that in coding regions (27.85%; Supplementary Table S3). Flower diameter had the highest number of associated SNPs (46.58% of the total significant SNPs), followed by plant size (21.46%) and maximum petal length (11.87%; Supplementary Figure S3, Supplementary Table S3). Moreover, the numbers of LD blocks of different lengths were summarized across the whole genome, and we identified that the length of most LD blocks was 50–100kb (Supplementary Figure S1F). Thus, we examined ± 50 kb regions surrounding each of the 70 representative SNPs; these regions

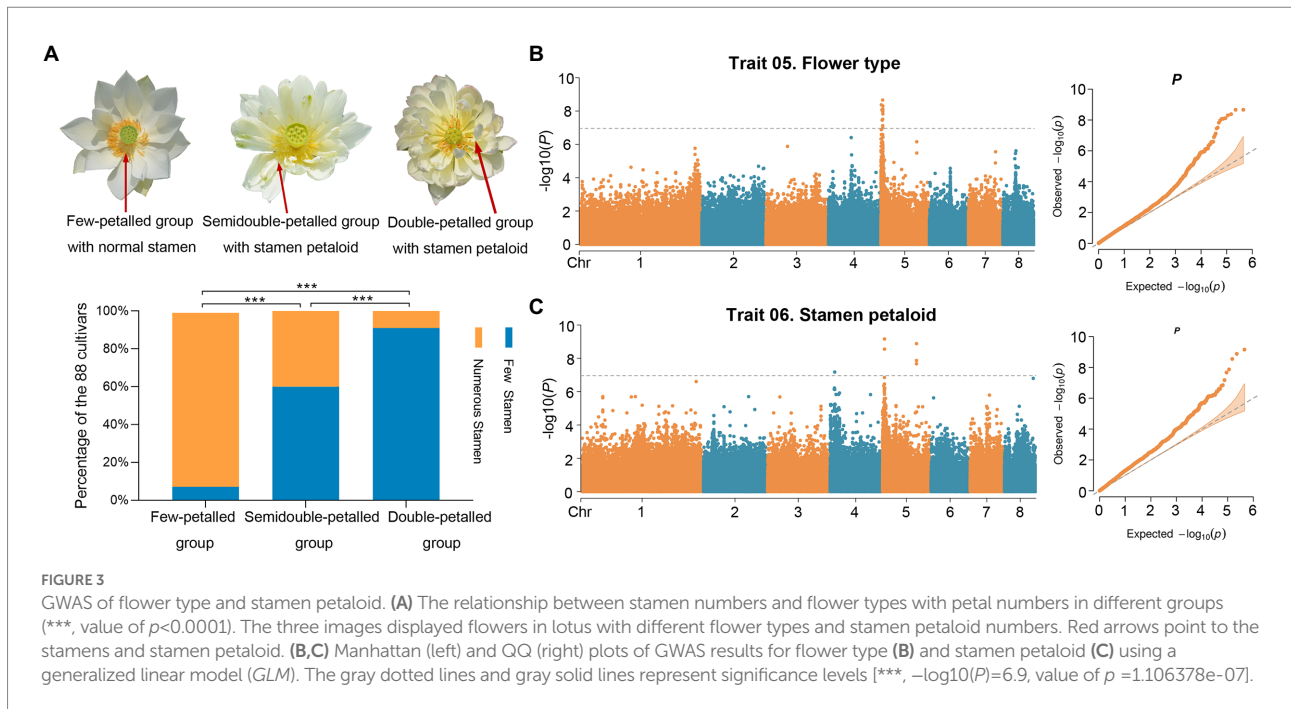
contained 149 candidate genes associated with phenotypes (Table 1; Supplementary Table S4). In the non-coding regions surrounding these candidate SNPs associated with the phenotypes, 28,764 associated non-coding RNAs (ncRNAs) and 6,317 TEs were found, which were frequently distributed on chromosome 1, encompassing 27.84% of the ncRNAs and 30.82% of the TEs (Figures 1D,F; Supplementary Tables S5, S6).

To gain insight into the functions of these trait-related candidate genes, we investigated the functional annotation of homologous genes in *Arabidopsis* and gene expression from various lotus tissues in the *Nelumbo Genome Database*. Among these genes, 25, 32 and 17 genes with known functions were distributed on chromosome 1, chromosome 4 and chromosome 7, respectively (Supplementary Figure S4). Further hierarchical clustering analysis of the gene expression profiles of various tissues for candidate genes can distinguish genes into several groups with modest yet consistent expression patterns, suggesting that such genes might share similar functions (Figure 4; Supplementary Figure S5; Eisen et al., 1998).

Candidate genes for flower types and stamen petaloid

Stamen petaloid and flower type, which mainly characterize petal layers and numbers, are two significantly correlated traits (Figure 2B). Notably, in these 88 lotus accessions, the number of stamens was inversely related to the number of petals through pairwise comparisons among the few-petalled lotus group, semidouble-petalled group and double-petalled group (χ^2 test of independence,





value of $p < 0.001$; Figure 3A). In the few-petalled lotus group, 92.86% had a few stamens, while the remaining flowers (7.14%) had numerous stamens; in the semidouble-petalled group, 60% of the accessions had a few stamens, while 40% had numerous stamens; and in the remaining groups with numerous petals, up to 86.49% exhibited few or extremely few stamens due to the petalised stamens. Our GWAS model uncovered five meaningful SNPs associated with 13 genes in three LD blocks for the two traits (Figures 3B,C; Table 1; Supplementary Table S4). The SNP locus (chr5_3,675,283) within the strong peak signal on chromosome 5 was associated with both stamen petaloid and flower type (Table 1; Figure 4A). To determine the expression differences in these genes between double-petalled and few-petalled flower groups, transcriptomes of lotus 'China Antique' (few-petalled) versus cultivated 'Xiaoxia' (double-petalled) at the formation period of flower buds and flower blooming stage were compared. Moreover, the differences in gene expression among petal, normal stamen, and stamen petaloid tissues in the same double-petalled individual, 'Fenhonglingxiao', were also compared to identify tissue-specific genes.

Among the five genes in the candidate region near this SNP locus (chr5_3,675,283), *OFPI17* (*Nn5g26879*) and *KUP2* (*Nn5g26880*) have defined orthologs in *Arabidopsis*, while *Nn5g26876*, *Nn5g26877* and *Nn5g26878* do not have any defined orthologs (Figure 4B; Table 1). *KUP2*, with a function in shoot cell expansion, was preferentially expressed in the flower apical buds of the few-petalled group (Figures 4C; Supplementary Figure S5A; Elumalai et al., 2002). However, in the double-petalled group, it was expressed the highest in stamen petaloid (Supplementary Figure S6A). The expression of *KUP2* in the apical buds and petals in the few-petalled lotus group was twofold that in the double-petalled lotus group, which might be considered that high expression of

KUP2 is likely required for proper petal development (Supplementary Figure S6B). The function of *Nn5g26877* was annotated as the transcriptional repressor *OFPI1*-like, which was highly expressed in immature and mature stamen tissues in the few-petalled group but barely expressed in stamen tissue in the double-petalled group, which suggested that its expression might promote the development of normal stamens in the few-petalled group, whereas the absence of expression in the double-petalled group might play a negative role in normal stamen development (Figure 4C; Supplementary Figures S5A, S6A). In line with the above conclusion, from apical buds to petals, the relative expression of this gene in the double-petalled group was threefold higher than that in the few-petalled group (Supplementary Figure S6B). These results implied that the expression of *KUP2* might contribute to stamen petaloid, while the expression of *Nn5g26877* might be involved in the development of normal stamens.

Among other associated genes for flower type and stamen petaloid, *PECT1* (*Nn5g26732*), which can modulate the interaction between phosphatidylcholine and *flowering locus T*, is expressed highest in flower apical buds in both the few-petalled and double-petalled groups (Figure 4C; Supplementary Figures S5; Susila et al., 2021). *Nn5g29478* was predicted to negatively regulate the formation of organ growth. It is expressed at the highest level in petals in both the few-petalled and double-petalled groups (Supplementary Figure S5A; Table 1). In the same few-petalled individual, *CML41* (*Nn5g29482*) expressed higher in stamen than petal tissues (Figure 4C; Supplementary Figure S4A). In comparison among stamens, stamen petaloid and petals within the same double-petalled individual, *CML41* expression was also the highest in normal stamens (Supplementary Figure S6A). However, when comparing the apical flower buds and fully

TABLE 1 The number of candidate coding genes significantly associated with flower types and stamen petaloid in 88 lotus accessions.

Trait	SNP Marker	P value	Lotus GeneID	Arabidopsis_ID (Name)	Function
Trait 05. Flower type	chr5-897,119	4.1826E-09	<i>Nn5g26732</i>	<i>AT2G38670 (PECT1)</i>	Phosphatidylethanolamine biosynthetic process.
	chr5-3,675,283	4.9692E-09	<i>Nn5g26876</i>	#N/A	#N/A
			<i>Nn5g26877</i>	#N/A	Transcription repressor <i>OFF1</i> -like, DNA-binding.
			<i>Nn5g26878</i>	#N/A	#N/A
			<i>Nn5g26879</i>	<i>AT2G30395 (OFF17)</i>	Member of the plant-specific ovate protein family.
			<i>Nn5g26880</i>	<i>AT2G40540 (KUP2)</i>	Putative potassium transporter <i>AtKT2p (AtKT2)</i> mRNA, potassium ion transport.
chr5-1,819,278	8.4980E-09	<i>Nn5g26787</i>	<i>AT5G01960</i>	Ubiquitin protein ligase activity.	
chr5-2,290,307	3.5133E-08	<i>Nn5g26803</i>	<i>AT5G58290 (RPT3)</i>	Positive regulation of RNA polymerase II transcription preinitiation complex assembly, protein catabolic process, ubiquitin-dependent protein catabolic process.	
Trait 06. Stamen petaloid	chr5-3,675,283	6.9103E-10	<i>Nn5g26876</i>	#N/A	#N/A
			<i>Nn5g26877</i>	#N/A	Transcription repressor <i>OFF1</i> -like, DNA-binding.
			<i>Nn5g26878</i>	#N/A	#N/A
			<i>Nn5g26879</i>	<i>AT2G30395 (OFF17)</i>	Member of the plant-specific ovate protein family.
			<i>Nn5g26880</i>	<i>AT2G40540 (KUP2)</i>	Putative potassium transporter <i>AtKT2p (AtKT2)</i> mRNA, Potassium ion transport.
Trait 06. Stamen petaloid	chr5-61,775,841	1.3503E-08	<i>Nn5g29478</i>	#N/A	Negative regulation of organ growth, ovule morphogenesis, seed morphogenesis, ubiquitin-binding, zinc ion binding.
			<i>Nn5g29482</i>	<i>AT3G50770 (CML41)</i>	Calcium-binding protein.
			<i>Nn5g29483</i>	#N/A	#N/A
			<i>Nn5g29484</i>	<i>AT4G34350 (HDR)</i>	Dimethylallyl diphosphate biosynthetic process, isopentenyl diphosphate biosynthetic process, methylerythritol 4-phosphate pathway.
			<i>Nn5g29485</i>	<i>AT3G47850</i>	<i>MYB</i> proto-oncogene protein, plant (A).

opened flowers between few-petal and double-petal individuals, *CML41* always had higher expression in double-petal accessions (Supplementary Figure S6B). Thus, the relatively higher expression of *CML41*, as a marker stamen-associated gene, in petals of double-petal compared with few-petal accessions might suggest that petals from double-petal flowers might originate from a mutated stamen.

Besides, we found a total of 2,100 small RNAs and 452 TEs associated with flower type and stamen petaloid, which might play roles in gene regulation (Figure 1F; Supplementary Tables S5, S6). Intriguingly, we found that stamen petaloid in double-petal flowers differed from that in few-petal flowers due to the deletion of a 2,755 bp DNA fragment (3,670,710–3,673,465) on chromosome 5 upstream of the *Nn5g26878* gene, which we named the lotus *Petaloid-Related Formation (PRF)* (Figure 4D). Within this region, there are seven TEs (three MITEs) and two small RNAs (19 nt- and 23 nt-sRNAs). In our study, all 14 few-petal lotus accessions, accounting for 15.91% of the 88 lotus accessions, retained the complete fragment in this region (χ^2 test of independence, value of $p < 0.001$). Approximately 85.94% of the accessions in the double-petal group completely lost this key genomic segment, which was significantly higher than that in both the few-petal (0%) and semidouble-petal (20.00%) lotus groups (χ^2 test of independence, value of $p < 0.001$; Figure 4E). In addition, to further confirm whether the deletion occurring in this

region in other lotus accessions is associated with petaloid stamens, we selected 136 accessions with sufficient read depth and a record of petal characteristics and uncovered a deletion in this region by BEDtools (Liu et al., 2020). In line with our result, 58.82% (10 of 17 double-petal lotuses) displayed complete deletions in the same region on chromosome 5, whereas 94.12% (112 of 119 few-petal lotuses) showed the complete presence of this genomic fragment (χ^2 test of independence, value of $p < 0.001$; Figure 4E). In parallel, we genotyped this presence and absence variant with Delly. The probability of missing genotypes within double-petal lotus (95.31%) was significantly higher than that within few-petal lotus (21.43%) and semidouble-petal lotus (0.00%) based on our 88 lotus accessions (χ^2 test of independence, value of $p < 0.001$; Supplementary Figure S7A). In the other 136 public accessions, the probability of missing genotypes with double-petal lotus was 58.82%, whereas that with few-petal lotus was only 2.52% (χ^2 test of independence, value of $p < 0.001$; Supplementary Figure S7B). Our investigation into the two independent datasets suggested that the absence of this region is likely responsible for double-petal flowers. Interestingly, the expression of *PRF*, which is most tightly linked to this region, was only found in the petals of double-petal flowers but not in those of few-petal flowers (Figure 4D). The relative expression of this gene in the petals of double-petal flowers was almost 30-fold that in few-petal flowers, suggesting that the expression of *PRF* can induce the

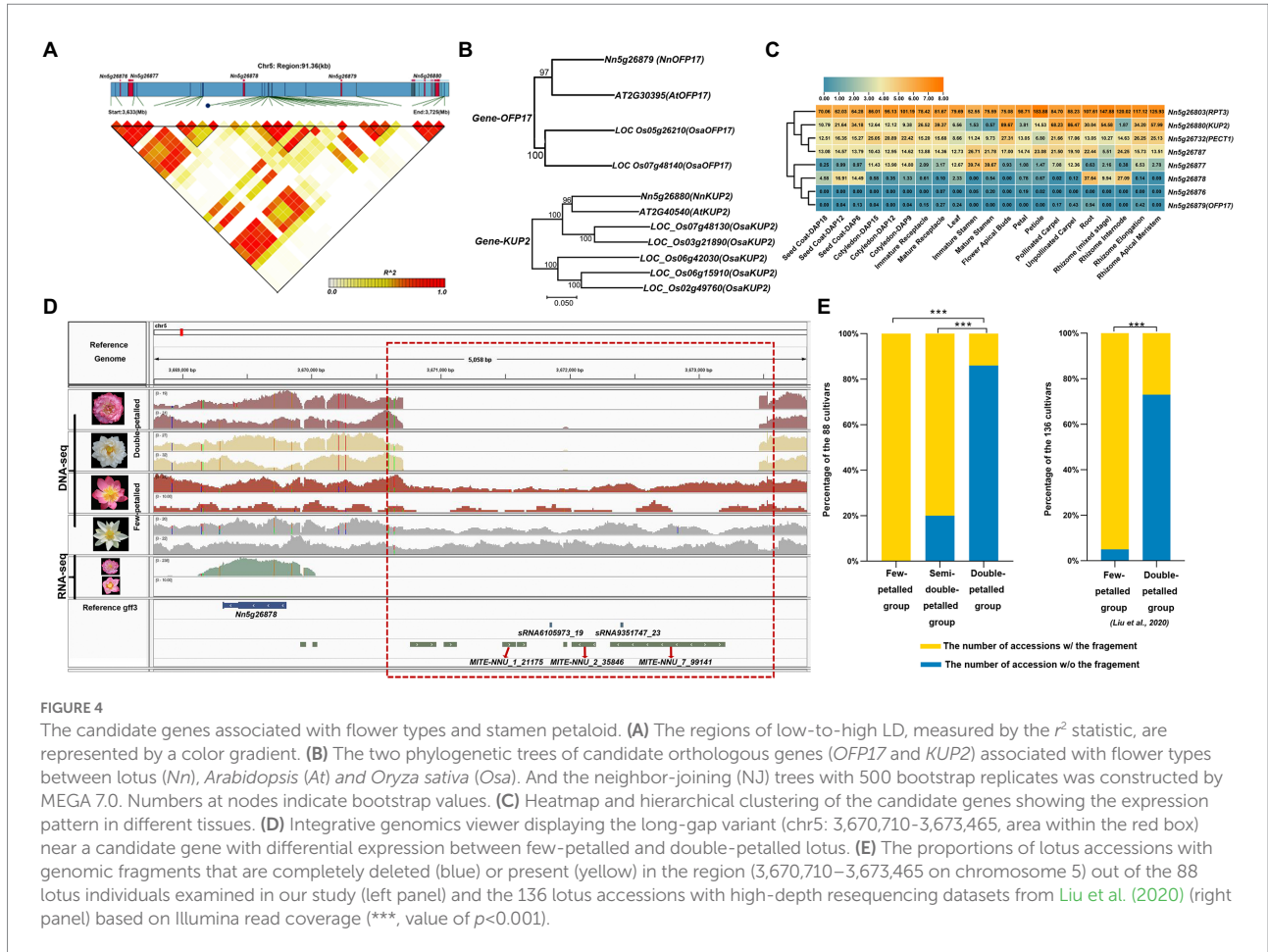


FIGURE 4

The candidate genes associated with flower types and stem petaloid. (A) The regions of low-to-high LD, measured by the r^2 statistic, are represented by a color gradient. (B) The two phylogenetic trees of candidate orthologous genes (*OFP17* and *KUP2*) associated with flower types between lotus (*Nn*), *Arabidopsis* (*At*) and *Oryza sativa* (*Osa*). And the neighbor-joining (NJ) trees with 500 bootstrap replicates was constructed by MEGA 7.0. Numbers at nodes indicate bootstrap values. (C) Heatmap and hierarchical clustering of the candidate genes showing the expression pattern in different tissues. (D) Integrative genomics viewer displaying the long-gap variant (chr5: 3,670,710–3,673,465, area within the red box) near a candidate gene with differential expression between few-petaloid and double-petaloid lotus. (E) The proportions of lotus accessions with genomic fragments that are completely deleted (blue) or present (yellow) in the region (3,670,710–3,673,465 on chromosome 5) out of the 88 lotus individuals examined in our study (left panel) and the 136 lotus accessions with high-depth resequencing datasets from Liu et al. (2020) (right panel) based on Illumina read coverage (***, value of $p < 0.001$).

formation of double-petaloid flowers (Supplementary Figure S6B). In the same double-petaloid cultivated lotus, the *PRF* gene was highly expressed in normally developed stamens (FPKM = 3.82) relative to its expression in the petaloid stamens (FPKM = 1.49) and petals (FPKM = 0.89) of this petaloid cultivar (Supplementary Figure S6A). These results collectively suggested that this *PRF* gene was abundantly expressed in double-petaloid flowers and likely contributed to petaloid formation.

Candidate genes for flower color and size variations

Four lotus genes were found in the candidate regions within LD blocks surrounding three significant SNPs based on GWAS. These genes were homologous to *Arabidopsis* *NOG1-1*, *TPPG*, *DJ1D* and *RABF2B* and were found to be associated with petal color variation (Supplementary Figures S3E, S5B; Supplementary Table S4). In *Arabidopsis*, these four genes play central roles in the ABA pathway, trehalose biosynthetic process, lactoylglutathione lyase activity, vacuole organization and transportation (Ueda et al., 2004; Ponnu et al., 2011; Kwon et al., 2013; Lee et al., 2017). Notably, the lotus *RABF2B* showed relatively high expression in flower apical buds and

petals of the red lotus, ‘China Antique’, which might help to transport anthocyanins to vacuoles (Supplementary Figure S5B). In addition to protein-coding genes with known functions, 2,192 ncRNAs and 556 TEs were also observed in the candidate region associated with flower coloration. One pre-miRNA, miR118, affects flower pigment accumulation through ATP sulfurylase (Glazinska et al., 2019; Supplementary Tables S5–S7).

Other than color, the shapes of lotus flowers are also diverse. However, only one candidate region was found by GWAS (Supplementary Figure S3F; Supplementary Table S4). In this candidate region, the gene *Nn5g27350* with an annotated function of RNA binding was found, which also exhibited the highest expression in apical buds (Supplementary Figures S3F, S5C; Supplementary Table S4).

In our study, pairwise Pearson correlations displayed relatively high correlation coefficients (r) among the three traits, flower diameter, petal length and width (Figure 2A), which was consistent with the development of typical flowers, such as *Petunia* (Cao et al., 2018). 36, 17 and 19 predicted genes with known functions were significantly associated with flower diameter, petal length and petal width, respectively, which were near 38 peak SNPs in candidate regions. Nine of these genes had detailed known functions associated with flower-related traits, such as *GRF2* (*Nn3g18854*)

with negative regulation of cell population proliferation and *NST1* (*Nn3g18857*) with functions in plant-type cell wall biogenesis (Supplementary Figures S3G–I; Supplementary Table S4; Mitsuda et al., 2005; Liu et al., 2012). Moreover, five genes, *CBSX3* (*Nn1g09382*), *LKR* (*Nn8g38508*), *CLC-B* (*Nn4g22356*), *ATL57* (*Nn2g13886*), and *ZIGA4* (*Nn2g14730*) were primarily expressed in flower apical buds or petals. These proteins have crucial functions in cell proliferation and apoptosis, lysine synthesis, intracellular vacuole structure, protein ubiquitination and plastid development, which may influence cell number and growth (Supplementary Figures S5D–F; Supplementary Table S4; Zhu and Galili, 2003; Koussevitzky et al., 2007; Jenny et al., 2010; Shin et al., 2020). Further investigation of the relative effects of these candidate genes on cell number or cell size on petal morphogenesis is needed.

Candidate genes associated with plant size variations

Plant size (architecture) is another crucial ornamental characteristic that is positively correlated with petal length, petal width and flower diameter (Figure 2B); plant size is determined by the coordinated progression of distinct organs based on the rate and duration of cell proliferation (Vercruyssen et al., 2011). A total of 54 genes located in 14 LD blocks that have known regulatory functions in the synthesis and transport of some proteins, cellular structure and jasmonate synthesis might be involved in the control of plant size in the development of lotus (Supplementary Figure S3J; Supplementary Table S4). *CLE26* (*Nn4g22977*), *ALA3* (*Nn4g22982*) and *VEL1* (*Nn8g40603*) were significantly associated with plant size variations involved in the regulation of cell differentiation in the phloem of roots and stems during auxin polar transport, vesicle-mediated transportation and the transition of the meristem from vegetative to reproductive phases (Supplementary Table S4; Kim and Sung, 2010; Czyzewicz and De Smet, 2016; Zhang X. et al., 2020). The *CLE26*, *ALA3* and *VEL1* genes were expressed primarily in the root apical meristem, immature stamen and root, respectively (Supplementary Figure S5G). This may be due to a close relationship between root growth and plant size. Notably, *ALA3* was likely a pleiotropic gene because it also had a significant effect on all three traits, petal length, petal width and flower diameter (Supplementary Table S3). *PRLIP* (*Nn3g19267*), belonging to the *PRLIP* family, which functions in the salicylic acid and ethylene signal transduction pathways, was expressed at the highest level in petioles (Jakab et al., 2003). Since petiole growth determines plant height, the petiole-expressed genes identified by GWAS might also be crucial in regulating plant size.

Inference of the functions of trait-related genes by coexpression network analysis

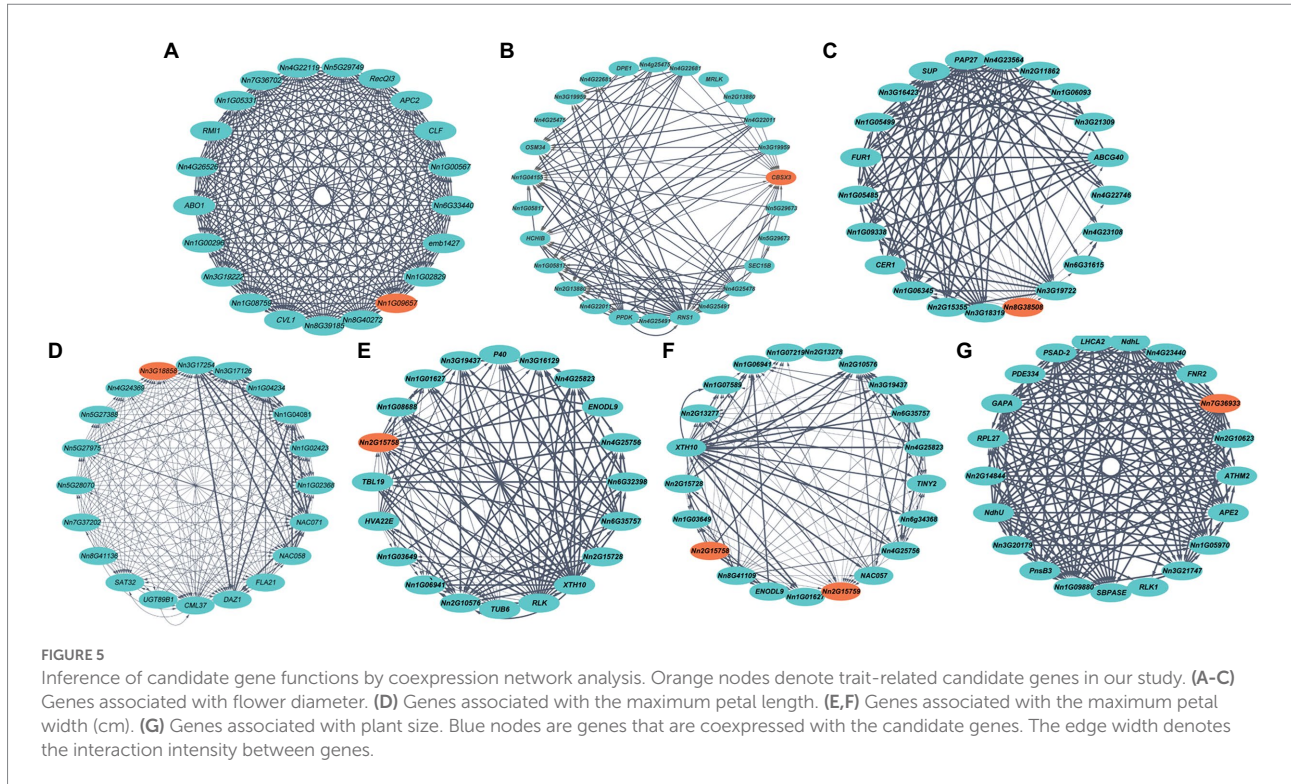
Gene coexpression networks based on diverse lotus tissues were downloaded from the *Nelumbo Genome Database* to identify genes

coexpressed with trait-related genes from our GWAS (Figure 5). Seven genes associated with flower diameter, maximum petal length or width, and plant size were found in the GCNs, suggesting that gene variants associated with plant growth were highly coexpressed as networks in lotus. Among them, although the *Nn1g09657* did not have a known homolog, it was annotated with a role in the mitochondrial organization. It is coexpressed with *ABO1* (*Nn2g12928*; Nelissen et al., 2010), *RMI1* (*Nn4g22247*; Knoll and Puchta, 2011), and *APC2* (*Nn4g23232*; Lorenzo-Orts et al., 2019), which are involved in cell division, differentiation and proliferation and likely regulates flower diameter (Figure 5A). The second petal diameter-associated candidate gene, *CBSX3* (*Nn1g09382*), is coexpressed with *SEC15B* (*Nn5g29674*; Fendrych et al., 2010), *MRLK* (*Nn5g29797*), *DPE1* (*Nn1g05190*; Yazdanbakhsh and Fisahn, 2011), and *HCHIB* (*Nn3g19481*; Vahabi et al., 2013), which have roles in positively regulating the growth of pollen tubes and meristematic cells and sugar catabolic processes (Figure 5B). One petal diameter and plant size candidate gene, *LKR* (*Nn8g38508*), is coexpressed with genes such as *Nn1g05485* and *ABCG40* (*Nn2g11872*; Kang et al., 2015), which have roles in the ethylene biosynthetic process, jasmonic acid biosynthetic process and ABA transport (Figure 5C) and likely contribute to plant hormones and growth. One maximum petal length candidate gene, *Nn3g18858*, is coexpressed with other genes involved in calcium channel and cytoskeleton formation, such as *CML37* (*Nn1g09382*; Scholz et al., 2014; Figure 5D). Two maximum petal width-related genes, *Nn2g15758* and *Nn2g15759*, are coexpressed with *Nn4g25756*, *TUB6* (*Nn4g24212*; Soga et al., 2018), *Nn6g35757*, and *XTH10* (*Nn6g34564*; Sasiharan et al., 2010) and function in cell wall formation (Figures 5E,F). One plant size-related gene, *MPH1* (*Nn7g36933*), was associated with 20 genes (such as *LHCA2*, *RPL27*, *RLK1*) in a network controlling the synthesis of chloroplast proteins; because chloroplasts are present in most of the skin cells, this result is in line with the fact that *MPH1* had the greatest expression in the leaves (Figure 5G; Liu and Last, 2015).

Discussion

GWAS of flower lotus

Finding the genetic basis of complex traits in plants, such as organ morphogenesis, has been a particular focus in the study of the evolution and domestication of flowering plants. In addition to coding regions, genetic variants in non-coding regions, have also attracted increasing attention (Chen, 2009; Axtell, 2013). As an important flower in water gardening that has been grown for thousands of years in East Asia, lotus exhibits a plethora of trait diversity, particularly flowers. In this study of a cultivated lotus population with morphological diversity, we not only identified a series of candidate genes significantly associated with various traits in coding regions but also highlighted the importance of small RNAs and TEs in non-coding regions based on GWAS, particularly for the petaloid and petal number traits.



GWAS has successfully identified a polymorphism in the gene *DWARF8* that is responsible for earlier flowering in maize (Thornsberry et al., 2001). Its effectiveness in identifying candidate genes for 107 phenotypes in *Arabidopsis* has also been demonstrated (Atwell et al., 2010). More studies using high-throughput sequencing techniques and computational models to investigate plants with complex traits with greatly accelerate direct discoveries for causal variants in GWAS according to genes located within LD blocks (Brachi et al., 2011; Gupta et al., 2019; Alseekh et al., 2021). The high-quality and representative genome assembly of well-studied species can also help us identify SNP loci and candidate genes for GWAS (Ji et al., 2015). In this study, the marker density of lotus at the genome level, with an average of one SNP per 0.56 kb, was already high enough to complete GWAS compared to the 0.5 SNPs/kb in 107 *Arabidopsis* accessions (Atwell et al., 2010), 1.2 SNPs/kb for 224 flax accessions (Xie et al., 2018) and 1.3 SNPs/kb for 419 cotton accessions (Ma et al., 2018). A prerequisite of GWAS is the comprehensive characterization of the LD pattern depending on the standing genetic diversity and the number of recombination events in the populations (Brachi et al., 2011). In *Arabidopsis*, the optimal number of SNPs was estimated to be 140,000 when LD was shown to decay within 10 kb (Kim et al., 2007). Correspondingly, the level of LD decayed within 32.7 kb, 455,292 SNPs were found in our study, which was almost consistent with the predicted necessary number of SNPs. At the methodological level, different SL-GWAS models, like *GLM*, *MLM* and *CMLM*, can also affect the effectiveness of GWAS. In the present study, significant associations between SNP loci and phenotypic characteristics were detected in *GLM* (Q) but not found in *MLM* or *CMLM* (Q+K), as we demonstrated that one

single population ($K = 1$) is optimal for our accessions. The core difference between *GLM*, *MLM* and *CMLM* is whether to control the kinship matrix K , and many studies have found that controlling K may lead to severe false negatives (Tibbs et al., 2021). To confirm the reliability of SNP loci in our study, the multi-locus GWAS (ML-GWAS) method with high detection power also needs to be considered, such as the study on the petal size of rapeseed (Qian et al., 2021). Generally, the number of significant SNPs associated with phenotypes by ML-GWAS was comparatively higher than those in the SL-GWAS model. A similar phenomenon could be found in the seed morphology of Brassica (280 SNP loci with ML-GWAS methods and 31 SNP loci with SL-GWAS methods; Khan et al., 2019) and plant architecture of hexaploid wheat (174 SNP loci with ML-GWAS methods, 97 SNP loci with SL-GWAS methods; Muhammad et al., 2021). This is likely because in ML-GWAS the effects of different loci in the whole genome can correct each other and then improve the detection efficiency of micro-effect loci. Also, this model generally does not form a series of significant SNPs, which will eliminate the collinearity caused by LD. Based on a suitable population, high marker density, and appropriate statistical models, GWAS has successfully explained the relationships between SNP markers and traits in other plants, such as inflorescence type in 82 *Hydrangea macrophylla* accessions (Wu and Alexander, 2020) and petal colors in 96 rose accessions (Schulz et al., 2016). In our study, we found a total of 149 possible candidate genes responsible for 12 traits, including flower development and plant size. There were also multiple high-quality SNPs in non-coding regions. Although only a few GWASs in plants have focused on non-coding regions with functional consequences associated with variants affecting the

expression of nearby genes, abundant variations in non-coding regions can accelerate the evolution of traits with weaker selective constraints (Giral et al., 2018; Broekema et al., 2020). In maize, three microRNAs have shown a significant association with leaf architecture (Tian et al., 2011). Therefore, one future focus on GWAS in lotus populations will be to verify the functions of key elements in non-coding regions that are associated with gene regulation through systematic transcriptomic and functional-genomic studies.

Unlike other traits in our study, no SNP locus or candidate genes were found to have significant correlations with flowering time, population floescence, flower density and mean petal number per flower for lotus. Therefore, more populations and phenotypic data on flowering time and density should be included to study these traits.

Molecular mechanisms underlying stamen petaloid formation

The normal formation of floral organs in each whorl of plants is directed by the classic ABCE model containing several *MADS*-box genes, which was discovered in *Arabidopsis* (Litt and Kramer, 2010). The aberrant development of the stamen in the second whorl leads to different flower types and has caused the evolution of lotus flower diversity, such as stamen petaloid. In previous studies, 11 homeotic *MADS*-box genes, one *AP2* gene and 31 methylation region-associated genes involved in stamen petaloid in lotus have been identified (Lin et al., 2018, 2019). Moreover, miR172, as a negative regulator, was also important during the crucial early stages of stamen development in lotus (Zhang Y. et al., 2020). However, the 16 candidate genes responsible for stamen petaloid variations in our study do not belong to the *MADS*-box family. They might be regulated in the downstream process by our GWAS-identified mutated genes responsible for flower type variations. Further molecular genetic studies are needed to investigate the regulatory relationship between our candidate genes and flower-type related *MADS*-box family.

Notably, we found novel structural variants in non-coding regions flanking these candidate genes responsible for stamen petaloid. The stamen petaloid in double-petaloid flowers could be caused by a deleted fragment in the non-coding region that includes seven TEs (three MITEs) and two small RNAs (sRNAs). TEs are classified into retrotransposons (class I) and DNA transposons (class II) based on their transposition mechanisms (Wicker et al., 2007). Among most experimental populations, the changes in the numbers of TEs, as the most variable parts of the genome, may lead to phenotypic differences even among closely related plant species because these elements can move to different locations (Lisch, 2013). It is widely recognized that TEs, particularly MITEs, play a key role in the evolution of phenotypic diversity from seed pigment to flowering time in maize (Selinger and Chandler, 1999; Salvi et al., 2007). In the Japanese morning glory, MITEs could insert into some flavonoid-related structural genes (*CHS*, *DFR* and *F3'5'H*) and transcription factors (*bHLH*), resulting in the loss of

pigment accumulation in flowers (Clegg and Durbin, 2000; Park et al., 2007), and an En/Spm-related TE inserted into the floral homeotic gene *DP* caused a double-flower phenotype (Nitasaka, 2003). MITEs have been suggested to be targets of small interfering RNAs (siRNAs) in rice, as a lower expression level was found in genes with MITE insertions (Lu et al., 2012). In our study, the gene expression level of *Nn5g26878* with MITE insertions in the promoter region was not detected in the few-petaloid flower, which was in line with the assumption that the MITE-associated gene silencing resulted in normal stamen development in lotus. Generally, most sRNAs perfectly matching MITEs belonged to the 21 nt class, which has been shown in wheat (Cantu et al., 2010). However, 19 nt- and 23 nt-sRNAs in this non-coding region were found. In addition to MITEs, there are five other TEs without detailed information in our study. Mature sRNAs, especially siRNAs as defense machinery, can silence the expression of transposable elements in the genome through RNA-directed DNA methylation, which has been reported in multiple organisms, including wheat (Cantu et al., 2010), rice (Nosaka et al., 2012) and *Drosophila* (Czech et al., 2008). Given that siRNA-binding MITEs could induce methylation levels in nearby regions, the deletion of siRNA-associated MITEs in the region near *PRF* might reduce methylation and consequently induce gene expression linked to the stamen petaloid phenotype. This variant has also been confirmed by our reanalysis of another independent lotus population dataset and can be widely applied in future genetic and breeding projects.

Data availability statement

These datasets presented in this study are deposited in the NCBI website (<https://www.ncbi.nlm.nih.gov/>). The accession numbers are SRP173547, SRP145546, SRP090666, PRJNA503979, PRJNA664744, and PRJNA417869.

Author contributions

TS, XY, and JC conceived the project and designed the experiments. ZG, YL, and YW collected samples for whole-genome resequencing. ZG performed the methodology and the formal analysis and wrote the original draft. TS and YX revised and edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by grants from the Strategic Priority Research Program of Chinese Academy of Sciences (No. XDB31000000), the Biological Resources Program, CAS (No. KFJ-BRP-007-009), the National Natural Science Foundation of China (Nos. 32170240, 31570220, and 31870208), the Youth Innovation Promotion Association of Chinese Academy of

Sciences (No. 2019335), and Bureau of Landscaping and Forestry of Wuhan Municipality (No. WHGF2019A10).

Acknowledgments

We thank Jia Li from Lanzhou University and Pingfang Yang from Hubei University for the discussion.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109
- Alseekh, S., Kostova, D., Bulut, M., and Fernie, A. R. (2021). Genome-wide association studies: assessing trait characteristics in model and crop plants. *Cell. Mol. Life Sci.* 78, 5743–5754. doi: 10.1007/s00018-021-03868-w
- Atwell, S., Huang, Y. S., Vilhjalmsón, B. J., Willems, G., Horton, M., Li, Y., et al. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465, 627–631. doi: 10.1038/nature08800
- Axtell, M. J. (2013). Classification and comparison of small RNAs from plants. *Annu. Rev. Plant Biol.* 64, 137–159. doi: 10.1146/annurev-arplant-050312-120043
- Brachi, B., Morris, G. P., and Borevitz, J. O. (2011). Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biol.* 12:232. doi: 10.1186/gb-2011-12-10-232
- Broekema, R. V., Bakker, O. B., and Jonkers, I. H. (2020). A practical view of fine-mapping and gene prioritization in the post-genome-wide association era. *Open Biol.* 10:190221. doi: 10.1098/rsob.190221
- Cantu, D., Vanzetti, L. S., Sumner, A., Dubcovsky, M., Matvienko, M., Distelfeld, A., et al. (2010). Small RNAs, DNA methylation and transposable elements in wheat. *BMC Genomics* 11:408. doi: 10.1186/1471-2164-11-408
- Cao, Z., Guo, Y., Yang, Q., He, Y., Fetouh, M. I., Warner, R. M., et al. (2018). Genome-wide search for quantitative trait loci controlling important plant and flower traits in petunia using an interspecific recombinant inbred population of *Petunia axillaris* and *Petunia exserta* G3: *Genes, Genomes, Genetics* 8, 2309–2317. doi: 10.1534/g3.118.200128
- Carroll, S. B. (2005). Evolution at two levels: on genes and form. *PLoS Biol.* 3:e245. doi: 10.1371/journal.pbio.0030245
- Castanera, R., Vendrell-Mir, P., Bardil, A., Carpentier, M. C., Panaud, O., and Casacuberta, J. M. (2021). Amplification dynamics of miniature inverted-repeat transposable elements and their impact on rice trait variability. *Plant J.* 107, 118–135. doi: 10.1111/tpj.15277
- Chen, X. (2009). Small RNAs and their roles in plant development. *Annu. Rev. Cell Dev. Biol.* 25, 21–44. doi: 10.1146/annurev.cellbio.042308.113417
- Clegg, M. T., and Durbin, M. L. (2000). Flower color variation: a model for the experimental study of evolution. *Proc. Natl. Acad. Sci. U. S. A.* 97, 7016–7023. doi: 10.1073/pnas.97.13.7016
- Czech, B., Malone, C. D., Zhou, R., Stark, A., Schlingehayde, C., Dus, M., et al. (2008). An endogenous small interfering RNA pathway in *Drosophila*. *Nature* 453, 798–802. doi: 10.1038/nature07007
- Czyzewicz, N., and De Smet, I. (2016). The *Arabidopsis thaliana* CLAVATA3/EMBRYO-SURROUNDING REGION 26 (CLE26) peptide is able to alter root architecture of *Solanum lycopersicum* and *Brassica napus*. *Plant Signal. Behav.* 11:e1118598. doi: 10.1080/15592324.2015.1118598
- Deng, J., Fu, Z., Chen, S., Damaris, R. N., Wang, K., Li, T., et al. (2015). Proteomic and epigenetic analyses of lotus (*Nelumbo nucifera*) petals between red and white cultivars. *Plant Cell Physiol.* 56, 1546–1555. doi: 10.1093/pcp/pcv077
- Deng, J., Li, J., Su, M., Lin, Z., Chen, L., and Yang, P. (2021). A bHLH gene *NiTT8* of *Nelumbo nucifera* regulates anthocyanin biosynthesis. *Plant Physiol. Biochem.* 158, 518–523. doi: 10.1016/j.plaphy.2020.11.038
- Diack, O., Kanfany, G., Gueye, M. C., Sy, O., Fofana, A., Tall, H., et al. (2020). GWAS unveils features between early- and late-flowering pearl millets. *BMC Genomics* 21:777. doi: 10.1186/s12864-020-07198-2
- Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A.* 95, 14863–14868. doi: 10.1073/pnas.95.25.14863
- Elumalai, R. P., Nagpal, P., and Reed, J. W. (2002). A mutation in the *Arabidopsis* *KT2/KUP2* potassium transporter gene affects shoot cell expansion. *Plant Cell* 14, 119–131. doi: 10.1105/tpc.010322
- Fendrych, M., Synek, L., Pecenkova, T., Toupalova, H., Cole, R., Drdova, E., et al. (2010). The *Arabidopsis* exocyst complex is involved in cytokinesis and cell plate maturation. *Plant Cell* 22, 3053–3065. doi: 10.1105/tpc.110.074351
- Gao, Z., Li, H., Yang, X., Yang, P., Chen, J., and Shi, T. (2021). Biased allelic expression in tissues of F1 hybrids between tropical and temperate lotus (*Nelumbo nucifera*). *Plant Mol. Biol.* 106, 207–220. doi: 10.1007/s11103-021-01138-8
- Giral, H., Landmesser, U., and Kratzer, A. (2018). Into the wild: GWAS exploration of non-coding RNAs. *Front Cardiovasc Med* 5:181. doi: 10.3389/fcvm.2018.00181
- Glazier, A. M., Nadeau, J. H., and Aitman, T. J. (2002). Finding genes that underlie complex traits. *Science* 298, 2345–2349. doi: 10.1126/science.1076641
- Glazinska, P., Kulasek, M., Glinkowski, W., Wojciechowski, W., and Kosinski, J. (2019). Integrated analysis of small RNA, transcriptome and degradome sequencing provides new insights into floral development and abscission in yellow lupine (*Lupinus luteus* L.). *Int. J. Mol. Sci.* 20:5122. doi: 10.3390/ijms20205122
- Guo, H. B. (2008). Cultivation of lotus (*Nelumbo nucifera* Gaertn. Ssp. *nucifera*) and its utilization in China. *Genet. Resour. Crop. Evol.* 56, 323–330. doi: 10.1007/s10722-008-9366-2
- Guo, S., Zhao, S., Sun, H., Wang, X., Wu, S., Lin, T., et al. (2019). Resequencing of 414 cultivated and wild watermelon accessions identifies selection for fruit quality traits. *Nat. Genet.* 51, 1616–1623. doi: 10.1038/s41588-019-0518-4
- Gupta, P. K., Kulwal, P. L., and Jaiswal, V. (2019). Association mapping in plants in the post-GWAS genomics era. *Adv. Genet.* 104, 75–154. doi: 10.1016/bs.adgen.2018.12.001
- Hayes, B. (2013). Overview of statistical methods for genome-wide association studies (GWAS). *Methods Mol. Biol.* 1019, 149–169. doi: 10.1007/978-1-62703-447-0_6
- Hoekstra, H. E., and Coyne, J. A. (2007). The locus of evolution: evo devo and the genetics of adaptation. *Evolution* 61, 995–1016. doi: 10.1111/j.1558-5646.2007.00105.x
- Huang, X., and Han, B. (2014). Natural variations and genome-wide association studies in crop plants. *Annu. Rev. Plant Biol.* 65, 531–551. doi: 10.1146/annurev-arplant-050213-035715
- Huang, L., Li, M., Cao, D., and Yang, P. (2021). Genetic dissection of rhizome yield-related traits in *Nelumbo nucifera* through genetic linkage map construction and QTL mapping. *Plant Physiol. Biochem.* 160, 155–165. doi: 10.1016/j.plaphy.2021.01.020
- Jakab, G., Manrique, A., Zimmerli, L., Metraux, J. P., and Mauch-Mani, B. (2003). Molecular characterization of a novel lipase-like pathogen-inducible gene family of *Arabidopsis*. *Plant Physiol.* 132, 2230–2239. doi: 10.1104/pp.103.025312

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.973347/full#supplementary-material>

- Jenny, V. D. F.-B., Martin, B., Marek, D., and Uwe, L. (2010). CLC-b-mediated NO-3/H+ exchange across the tonoplast of *Arabidopsis* vacuoles. *Plant Cell Physiol.* 51, 960–968. doi: 10.1093/pcp/pcq062
- Ji, L., Neumann, D. A., and Schmitz, R. J. (2015). Crop epigenomics: identifying, unlocking, and harnessing cryptic variation in crop genomes. *Mol. Plant* 8, 860–870. doi: 10.1016/j.molp.2015.01.021
- Jiao, Y., Zhao, H., Ren, L., Song, W., Zeng, B., Guo, J., et al. (2012). Genome-wide genetic changes during modern breeding of maize. *Nat. Genet.* 44, 812–815. doi: 10.1038/ng.2312
- Kang, J., Yim, S., Choi, H., Kim, A., Lee, K. P., Lopez-Molina, L., et al. (2015). Abscisic acid transporters cooperate to control seed germination. *Nat. Commun.* 6:8113. doi: 10.1038/ncomms9113
- Khan, S. U., Yangmiaou, J., Liu, S., Zhang, K., Khan, M. H. U., Zhai, Y., et al. (2019). Genome-wide association studies in the genetic dissection of ovule number, seed number, and seed weight in *Brassica napus* L. *Ind. Crop. Prod.* 142:111877. doi: 10.1016/j.indcrop.2019.111877
- Kim, S., Plagnol, V., Hu, T. T., Toomajian, C., Clark, R. M., Ossowski, S., et al. (2007). Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nat. Genet.* 39, 1151–1155. doi: 10.1038/ng2115
- Kim, D. H., and Sung, S. (2010). The plant Homeo domain finger protein, VIN3-LIKE 2, is necessary for photoperiod-mediated epigenetic regulation of the floral repressor, MAF5. *Proc. Natl. Acad. Sci. U. S. A.* 107, 17029–17034. doi: 10.1073/pnas.1010834107
- Knoll, A., and Puchta, H. (2011). The role of DNA helicases and their interaction partners in genome stability and meiotic recombination in plants. *J. Exp. Bot.* 62, 1565–1579. doi: 10.1093/jxb/erq357
- Koussevitzky, S., Stanne, T. M., Peto, C. A., Giap, T., Sjogren, L. L., Zhao, Y., et al. (2007). An *Arabidopsis thaliana* virescent mutant reveals a role for *ClpR1* in plastid development. *Plant Mol. Biol.* 63, 85–96. doi: 10.1007/s1103-006-9074-2
- Kwon, K., Choi, D., Hyun, J. K., Jung, H. S., Baek, K., and Park, C. (2013). Novel glyoxalases from *Arabidopsis thaliana*. *FEBS J.* 280, 3328–3339. doi: 10.1111/febs.12321
- Lee, T. H., Guo, H., Wang, X., Kim, C., and Paterson, A. H. (2014). SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics* 15:162. doi: 10.1186/1471-2164-15-162
- Lee, S., Senthil-Kumar, M., Kang, M., Rojas, C. M., Tang, Y., Oh, S., et al. (2017). The small GTPase, nucleolar GTP-binding protein 1 (NOG1), has a novel role in plant innate immunity. *Sci. Rep.* 7:9260. doi: 10.1038/s41598-017-08932-9
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv [preprint], 1303.3997. doi: 10.48550/arXiv.1303.3997
- Li, J., Shi, T., Huang, L., He, D., Nyong, A. T. M., and Yang, P. (2018). Systematic transcriptomic analysis provides insights into lotus (*Nelumbo nucifera*) seed development. *Plant Growth Regul.* 86, 339–350. doi: 10.1007/s10725-018-0433-1
- Li, H., Yang, X., Wang, Q., Chen, J., and Shi, T. (2021a). Distinct methylome patterns contribute to ecotypic differentiation in the growth of the storage organ of a flowering plant (sacred lotus). *Mol. Ecol.* 30, 2831–2845. doi: 10.1111/mec.15933
- Li, H., Yang, X., Zhang, Y., Gao, Z., Liang, Y., Chen, J., et al. (2021b). *Nelumbo genome database*, an integrative resource for gene expression and variants of *Nelumbo nucifera*. *Sci Data* 8:38. doi: 10.1038/s41597-021-00828-8
- Lin, Z., Damaris, R. N., Shi, T., Li, J., and Yang, P. (2018). Transcriptomic analysis identifies the key genes involved in stamen petaloid in lotus (*Nelumbo nucifera*). *BMC Genomics* 19:554. doi: 10.1186/s12864-018-4950-0
- Lin, Z., Liu, M., Damaris, R. N., Nyong'a, T. M., Cao, D., Ou, K., et al. (2019). Genome-wide DNA methylation profiling in the lotus (*Nelumbo nucifera*) flower showing its contribution to the stamen petaloid. *Plants (Basel)* 8:135. doi: 10.3390/plants8050135
- Lin, T., Zhu, G., Zhang, J., Xu, X., Yu, Q., Zheng, Z., et al. (2014). Genomic analyses provide insights into the history of tomato breeding. *Nat. Genet.* 46, 1220–1226. doi: 10.1038/ng.3117
- Lisch, D. (2013). How important are transposons for plant evolution? *Nat. Rev. Genet.* 14, 49–61. doi: 10.1038/nrg3374
- Litt, A., and Kramer, E. M. (2010). The ABC model and the diversification of floral organ identity. *Semin. Cell Dev. Biol.* 21, 129–137. doi: 10.1016/j.semcdb.2009.11.019
- Liu, J., Hua, W., Yang, H. L., Zhan, G. M., Li, R. J., Deng, L. B., et al. (2012). The BnGRF2 gene (GRF2-like gene from *Brassica napus*) enhances seed oil production through regulating cell number and plant photosynthesis. *J. Exp. Bot.* 63, 3727–3740. doi: 10.1093/jxb/ers066
- Liu, J., and Last, R. L. (2015). A land plant-specific thylakoid membrane protein contributes to photosystem II maintenance in *Arabidopsis thaliana*. *Plant J.* 82, 731–743. doi: 10.1111/tpj.12845
- Liu, Z., Zhu, H., Zhou, J., Jiang, S., Wang, Y., Kuang, J., et al. (2020). Resequencing of 296 cultivated and wild lotus accessions unravels its evolution and breeding history. *Plant J.* 104, 1673–1684. doi: 10.1111/tpj.15029
- Lorenzo-Orts, L., Withthoeft, J., Deforges, J., Martinez, J., Loubery, S., Placzek, A., et al. (2019). Concerted expression of a cell cycle regulator and a metabolic enzyme from a bicistronic transcript in plants. *Nat. Plants* 5, 184–193. doi: 10.1038/s41477-019-0358-3
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8
- Lu, C., Chen, J., Zhang, Y., Hu, Q., Su, W., and Kuang, H. (2012). Miniature inverted-repeat transposable elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in *Oryza sativa*. *Mol. Biol. Evol.* 29, 1005–1017. doi: 10.1093/molbev/msr282
- Ma, Z., He, S., Wang, X., Sun, J., Zhang, Y., Zhang, G., et al. (2018). Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. *Nat. Genet.* 50, 803–813. doi: 10.1038/s41588-018-0119-7
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- Ming, R., VanBuren, R., Liu, Y., Yang, M., Han, Y., Li, L. T., et al. (2013). Genome of the long-living sacred lotus (*Nelumbo nucifera* Gaertn.). *Genome Biol.* 14:R41. doi: 10.1186/gb-2013-14-5-r41
- Mitsuda, N., Seki, M., Shinozaki, K., and Ohme-Takagi, M. (2005). The NAC transcription factors *NST1* and *NST2* of *Arabidopsis* regulate secondary wall thickenings and are required for anther dehiscence. *Plant Cell* 17, 2993–3006. doi: 10.1105/tpc.105.036004
- Muhammad, A., Li, J., Hu, W., Yu, J., Khan, S. U., Khan, M. H. U., et al. (2021). Uncovering genomic regions controlling plant architectural traits in hexaploid wheat using different GWAS models. *Sci. Rep.* 11:6767. doi: 10.1038/s41598-021-86127-z
- Nelissen, H., De Groeve, S., Fleury, D., Neyt, P., Bruno, L., Bitonti, M. B., et al. (2010). Plant elongator regulates auxin-related genes during RNA polymerase II transcription elongation. *Proc. Natl. Acad. Sci. U. S. A.* 107, 1678–1683. doi: 10.1073/pnas.0913559107
- Nitasaka, E. (2003). Insertion of an en/Spm-related transposable element into a floral homeotic gene *DUPLICATED* causes a double flower phenotype in the Japanese morning glory. *Plant J.* 36, 522–531. doi: 10.1046/j.1365-3113.2003.01896.x
- Nosaka, M., Itoh, J., Nagato, Y., Ono, A., Ishiwata, A., and Sato, Y. (2012). Role of transposon-derived small RNAs in the interplay between genomes and parasitic DNA in rice. *PLoS Genet.* 8:e1002953. doi: 10.1371/journal.pgen.1002953
- Parachnowitsch, A. L., and Kessler, A. (2010). Pollinators exert natural selection on flower size and floral display in *Penstemon digitalis*. *New Phytol.* 188, 393–402. doi: 10.1111/j.1469-8137.2010.03410.x
- Park, K. I., Ishikawa, N., Morita, Y., Choi, J. D., Hoshino, A., and Iida, S. (2007). A *bHLH* regulatory gene in the common morning glory, *Ipomoea purpurea*, controls anthocyanin biosynthesis in flowers, proanthocyanidin and phytomelanin pigmentation in seeds, and seed trichome formation. *Plant J.* 49, 641–654. doi: 10.1111/j.1365-3113.2006.02988.x
- Payne, J. L., and Wagner, A. (2019). The causes of evolvability and their evolution. *Nat. Rev. Genet.* 20, 24–38. doi: 10.1038/s41576-018-0069-z
- Perteau, M., Kim, D., Perteau, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* 11, 1650–1667. doi: 10.1038/nprot.2016.095
- Ponnu, J., Wahl, V., and Schmid, M. (2011). Trehalose-6-phosphate: connecting plant metabolism and development. *Front. Plant Sci.* 2:70. doi: 10.3389/fpls.2011.00070
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Qian, M., Fan, Y., Li, Y., Liu, M., Sun, W., Duan, H., et al. (2021). Genome-wide association study and transcriptome comparison reveal novel QTL and candidate genes that control petal size in rapeseed. *J. Exp. Bot.* 72, 3597–3610. doi: 10.1093/jxb/erab105
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033
- Rausch, T., Zichner, T., Schlattl, A., Stutz, A. M., Benes, V., and Korbel, J. O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28, i333–i339. doi: 10.1093/bioinformatics/bts378
- Romero, I. G., Ruvinsky, I., and Gilad, Y. (2012). Comparative studies of gene expression and the evolution of gene regulation. *Nat. Rev. Genet.* 13, 505–516. doi: 10.1038/nrg3229
- Ruiz-Sola, M. A., and Rodriguez-Concepcion, M. (2012). Carotenoid biosynthesis in *Arabidopsis*: a colorful pathway. *Arabidopsis Book* 10:e0158. doi: 10.1199/tab.0158

- Saito, K., Yonekura-Sakakibara, K., Nakabayashi, R., Higashi, Y., Yamazaki, M., Tohge, T., et al. (2013). The flavonoid biosynthetic pathway in Arabidopsis: structural and genetic diversity. *Plant Physiol. Biochem.* 72, 21–34. doi: 10.1016/j.plaphy.2013.02.001
- Salvi, S., Sponza, G., Morgante, M., Tomes, D., Niu, X., Fengler, K. A., et al. (2007). Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc. Natl. Acad. Sci. U. S. A.* 104, 11376–11381. doi: 10.1073/pnas.0704145104
- Sasidharan, R., Chinnappa, C. C., Staal, M., Elzenga, J. T., Yokoyama, R., Nishitani, K., et al. (2010). Light quality-mediated petiole elongation in *Arabidopsis* during shade avoidance involves cell wall modification by xyloglucan endotransglucosylase/hydrolases. *Plant Physiol.* 154, 978–990. doi: 10.1104/pp.110.162057
- Scholz, S. S., Vadassery, J., Heyer, M., Reichelt, M., Bender, K. W., Snedden, W. A., et al. (2014). Mutation of the *Arabidopsis* calmodulin-like protein CML37 deregulates the jasmonate pathway and enhances susceptibility to herbivory. *Mol. Plant* 7, 1712–1726. doi: 10.1093/mp/ssu102
- Schulz, D. F., Schott, R. T., Voorrips, R. E., Smulders, M. J., Linde, M., and Debener, T. (2016). Genome-wide association analysis of the anthocyanin and carotenoid contents of rose petals. *Front. Plant Sci.* 7:1798. doi: 10.3389/fpls.2016.01798
- Selinger, D. A., and Chandler, V. L. (1999). Major recent and independent changes in levels and patterns of expression have occurred at the b gene, a regulatory locus in maize. *Proc. Natl. Acad. Sci. U. S. A.* 96, 15007–15012. doi: 10.1073/pnas.96.26.15007
- Shastry, B. S. (2009). SNPs: impact on gene function and phenotype. *Methods Mol. Biol.* 578, 3–22. doi: 10.1007/978-1-60327-411-1_1
- Shi, T., Rahmani, R. S., Gugger, P. F., Wang, M., Li, H., Zhang, Y., et al. (2020). Distinct expression and methylation patterns for genes with different fates following a single whole-genome duplication in flowering plants. *Mol. Biol. Evol.* 37, 2394–2413. doi: 10.1093/molbev/msaa105
- Shi, T., Wang, K., and Yang, P. (2017). The evolution of plant microRNAs: insights from a basal eudicot sacred lotus. *Plant J.* 89, 442–457. doi: 10.1111/tpj.13394
- Shin, J. S., So, W. M., Kim, S. Y., Noh, M., Hyoung, S., Yoo, K. S., et al. (2020). CBSX3-Trxo-2 regulates ROS generation of mitochondrial complex II (succinate dehydrogenase) in Arabidopsis. *Plant Sci.* 294:110458. doi: 10.1016/j.plantsci.2020.110458
- Signor, S. A., and Nuzhdin, S. V. (2018). The evolution of gene expression in *cis* and *trans*. *Trends Genet.* 34, 532–544. doi: 10.1016/j.tig.2018.03.007
- Soga, K., Yamazaki, C., Kamada, M., Tanigawa, N., Kasahara, H., Yano, S., et al. (2018). Modification of growth anisotropy and cortical microtubule dynamics in *Arabidopsis* hypocotyls grown under microgravity conditions in space. *Physiol. Plant.* 162, 135–144. doi: 10.1111/pp.12640
- Song, J. M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., et al. (2020). Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nat Plants* 6, 34–45. doi: 10.1038/s41477-019-0577-7
- Sun, S. S., Gugger, P. F., Wang, Q. F., and Chen, J. M. (2016). Identification of a *R2R3-MYB* gene regulating anthocyanin biosynthesis and relationships between its variation and flower color difference in lotus (*Nelumbo Adans.*). *PeerJ* 4:e2369. doi: 10.7717/peerj.2369
- Susila, H., Nasim, Z., Gawarecka, K., Jung, J. Y., Jin, S., Youn, G., et al. (2021). *PHOSPHORYLETHANOLAMINE CYTIDYLTRANSFERASE 1* modulates flowering in a florigen-independent manner by regulating SVP. *Development* 148:870. doi: 10.1242/dev.193870
- Syvanen, A. C. (2001). Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat. Rev. Genet.* 2, 930–942. doi: 10.1038/35103535
- Takeda, S., Iwasaki, A., Matsumoto, N., Uemura, T., Tatematsu, K., and Okada, K. (2013). Physical interaction of floral organs controls petal morphogenesis in *Arabidopsis*. *Plant Physiol.* 161, 1242–1250. doi: 10.1104/pp.112.212084
- Thornsberry, J. M., Goodman, M. M., Doebley, J., Kresovich, S., Nielsen, D., and Buckler, E. S. T. (2001). *Dwarf8* polymorphisms associate with variation in flowering time. *Nat. Genet.* 28, 286–289. doi: 10.1038/90135
- Tian, F., Bradbury, P. J., Brown, P. J., Hung, H., Sun, Q., Flint-Garcia, S., et al. (2011). Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat. Genet.* 43, 159–162. doi: 10.1038/ng.746
- Tibbs, C. L., Zhang, Z., and Yu, J. (2021). Status and prospects of genome-wide association studies in plants. *Plant Genome* 14:e20077. doi: 10.1002/tpg2.20077
- Tornielli, G., Koes, R., and Quattrocchio, F. (2009). The genetics of flower color. *Petunia* 107, 269–299. doi: 10.1007/978-0-387-84796-2_13
- Ueda, T., Uemura, T., Sato, M. H., and Nakano, A. (2004). Functional differentiation of endosomes in *Arabidopsis* cells. *Plant J.* 40, 783–789. doi: 10.1111/j.1365-3113X.2004.02249.x
- Vahabi, K., Camehl, I., Sherameti, I., and Oelmüller, R. (2013). Growth of *Arabidopsis* seedlings on high fungal doses of *Piriformospora indica* has little effect on plant performance, stress, and defense gene expression in spite of elevated jasmonic acid and jasmonic acid-isoleucine levels in the roots. *Plant Signal. Behav.* 8:e26301. doi: 10.4161/psb.26301
- Vercruyssen, L., Gonzalez, N., Werner, T., Schumling, T., and Inze, D. (2011). Combining enhanced root and shoot growth reveals cross talk between pathways that control plant organ size in *Arabidopsis*. *Plant Physiol.* 155, 1339–1352. doi: 10.1104/pp.110.167049
- Wang, Q. (2005). *Lotus flower cultivars in China*. China Forestry Publishing House: Beijing, China.
- Wang, Y., Chen, Y., Yuan, M., Xue, Z., Jin, Q., and Xu, Y. (2016). Flower color diversity revealed by differential expression of flavonoid biosynthetic genes in sacred lotus. *J. Am. Soc. Hortic. Sci.* 141, 573–582. doi: 10.21273/jashs03848-16
- Wang, J., Li, L., Li, C., Yang, X., Xue, Y., Zhu, Z., et al. (2021). A transposon in the vacuolar sorting receptor gene *TaVSR1-B* promoter region is associated with wheat root depth at booting stage. *Plant Biotechnol. J.* 19, 1456–1467. doi: 10.1111/pbi.13564
- Wang, Q., and Zhang, X. (2005). *Colored illustration of lotus cultivars in China*. China Forestry Publishing House: Beijing, China.
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., et al. (2007). A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8, 973–982. doi: 10.1038/nrg2165
- Wu, X., and Alexander, L. W. (2020). Genome-wide association studies for inflorescence type and remontancy in *Hydrangea macrophylla*. *Hortic Res* 7:27. doi: 10.1038/s41438-020-0255-y
- Xie, D., Dai, Z., Yang, Z., Tang, Q., Sun, J., Yang, X., et al. (2018). Genomic variations and association study of agronomic traits in flax. *BMC Genomics* 19:512. doi: 10.1186/s12864-018-4899-z
- Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82. doi: 10.1016/j.ajhg.2010.11.011
- Yang, M., Zhu, L., Pan, C., Xu, L., Liu, Y., Ke, W., et al. (2015). Transcriptomic analysis of the regulation of rhizome formation in temperate and tropical lotus (*Nelumbo nucifera*). *Sci. Rep.* 5:13059. doi: 10.1038/srep13059
- Yang, M., Zhu, L., Xu, L., Pan, C., and Liu, Y. (2014). Comparative transcriptomic analysis of the regulation of flowering in temperate and tropical lotus (*Nelumbo nucifera*) by RNA-Seq. *Ann. Appl. Biol.* 165, 73–95. doi: 10.1111/aab.12119
- Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P. C., Hu, L., et al. (2016). Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat. Genet.* 48, 927–934. doi: 10.1038/ng.3596
- Yazdanbakhsh, N., and Fisahn, J. (2011). Mutations in leaf starch metabolism modulate the diurnal root growth profiles of *Arabidopsis thaliana*. *Plant Signal. Behav.* 6, 995–998. doi: 10.4161/psb.6.7.15484
- Zhang, X., Adamowski, M., Marhava, P., Tan, S., Zhang, Y., Rodriguez, L., et al. (2020). *Arabidopsis* flippases cooperate with ARF GTPase exchange factors to regulate the trafficking and polarity of PIN auxin transporters. *Plant Cell* 32, 1644–1664. doi: 10.1105/tpc.19.00869
- Zhang, C., Dong, S. S., Xu, J. Y., He, W. M., and Yang, T. L. (2019). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35, 1786–1788. doi: 10.1093/bioinformatics/bty875
- Zhang, Y., Rahmani, R. S., Yang, X., Chen, J., and Shi, T. (2020). Integrative expression network analysis of microRNA and gene isoforms in sacred lotus. *BMC Genomics* 21:429. doi: 10.1186/s12864-020-06853-y
- Zhang, J., Song, Q., Cregan, P. B., Nelson, R. L., Wang, X., Wu, J., et al. (2015). Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. *BMC Genomics* 16:217. doi: 10.1186/s12864-015-1441-4
- Zhang, X. Y., and Wang, Q. C. (2011). *New lotus flower cultivars in China I*. China Forestry Publishing House: Beijing, China.
- Zhang, L., Zhang, F., Liu, F., Shen, J., Wang, J., Jiang, M., et al. (2021). The lotus *NnFTIP1* and *NnFTI* regulate flowering time in Arabidopsis. *Plant Sci.* 302:110677. doi: 10.1016/j.plantsci.2020.110677
- Zhang, Q., Zhang, H., Sun, L., Fan, G., Ye, M., Jiang, L., et al. (2018). The genetic architecture of floral traits in the woody plant *Prunus mume*. *Nat. Commun.* 9:1702. doi: 10.1038/s41467-018-04093-z
- Zhu, X., and Galili, G. (2003). Increased lysine synthesis coupled with a knockout of its catabolism synergistically boosts lysine content and also transregulates the metabolism of other amino acids in *Arabidopsis* seeds. *Plant Cell* 15, 845–853. doi: 10.1105/tpc.009647
- Zhu, H. H., Yang, J. X., Xiao, C. H., Mao, T. Y., Zhang, J., and Zhang, H. Y. (2019). Differences in flavonoid pathway metabolites and transcripts affect yellow petal colouration in the aquatic plant *Nelumbo nucifera*. *BMC Plant Biol.* 19:277. doi: 10.1186/s12870-019-1886-8