Selection pressure causes differentiation of the *SPL* gene family in the Juglandaceae

Di Wu¹ · Jia Luo¹ · Junhao Chen¹ · Liangsheng Zhang² · Kean-Jin Lim¹ · Zhengjia Wang¹

Received: 3 August 2018 / Accepted: 3 April 2019 / Published online: 11 April 2019
© The Author(s) 2019

Abstract

The SQUAMOSA promoter-binding protein-like (*SPL*) is a plant-specific transcription factor that influences flowering and vegetative development. Although the *SPL* genes have been functionally analyzed in many species, studies on the evolutionary history of the whole gene family, and in the Juglandaceae specifically, have been limited. Here, we conducted a phylogenetic relationship analysis of the Juglandaceae *SPL* gene family compared with other land plant species. Our results showed that the *SPL* genes were divided into three major clades, all of which were further divided into ten small clades. Selection pressure analysis suggested that all *SPL* genes were exposed to purifying selection pressure during evolution. The purifying selection was smaller for the Juglandaceae *SPL* genes than for other angiosperms, indicating a greater susceptibility to functional differentiation in the Juglandaceae. The *SPL* proteins encoded by Clade 1 contained a branch-specific transmembrane structure and many conserved motif combinations at the C-terminal. We also detected many selection sites in these motif combinations. Expression analysis showed that Clade 1 genes had spatial and temporal differences and were highly expressed in various organs. The expression profile was closely related to the selection sites and motif combinations at the C-terminal. These observations represent essential entry points for revealing the functional differentiation of the *SPL* gene family. Our data presented here may provide a basis for future investigations of *SPL* genes in the Juglandaceae, especially for flower development and perhaps crop yield improvement.

Keywords *SPL* · Juglandaceae · Phylogenetic analysis · Selection pressure

Introduction

Transcription factors are vital for plant growth and development because they regulate numerous physiological and biochemical processes. They can promote cell division or prevent an early division of immature cells (Graaff et al. 2009), thereby regulating the development of inflorescence

meristems, ovules, seeds, and flower organs (Riechmann and Meyerowitz 1998). Flower and vegetative development is regulated by the plant-specific SQUAMOSA promoter-binding protein-like (*SPL*) transcription factor family (Wang et al. 2009). The *SPL* proteins can be separated into two major groups—the long and short proteins—and the short protein group is regulated by the microRNAs *miR156* and *miR157* (Xing et al. 2010). The long *SPL* proteins comprise more than 800 amino acids, while the short *SPL* proteins contain no more than 400 amino acids (Xing et al. 2010). These differences are responsible for the considerable functional diversity among these transcription factors, which influence plant morphology, flower and fruit development, and transduction of hormone signals, as well as other physiological processes (Gou et al. 2011). All *SPL* proteins contain a conserved SQUAMOSA PROMOTER-BINDING PROTEIN (SBP) domain entailing about 78 amino acids (Cardon et al. 1999; Birkenbihl et al. 2005; Guo et al. 2008). This domain contains three functionally important sequences: zinc finger 1 (Zn1), zinc finger 2 (Zn2) (Yamasaki et al.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00438-019-01562-y>) contains supplementary material, which is available to authorized users.

✉ Kean-Jin Lim
keanjin.lim@zafu.edu.cn

✉ Zhengjia Wang
wzhj21@163.com

¹ State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University, No. 666 Wusu Road, Lin'an District, Hangzhou 311300, Zhejiang, China

² Center for Genomics and Biotechnology, Fujian Agriculture and Forestry University, Fuzhou, Fujian, China

2004), and a nuclear location signal (NLS) (Birkenbihl et al. 2005). The SBP domain binds to the consensus nucleotide sequence TNCGTACAA of the *SQUAMOSA* promoter to regulate the expression of the *SPL* gene (Cardon et al. 1997).

The *SPL* genes are expressed in different plant tissues during different growth stages, but they are mainly expressed in young vegetative and flowering organ tissues. They have been partitioned by differential subfunctionalization following gene duplication and speciation, which have indicated that neo-functionalization and parallel recruitment occurred relatively recently in evolution (Preston and Hileman 2013). Thus, the numbers of *SPL* gene family members differ within plant species; for example, 16 *SPL* genes are reported in *Arabidopsis thaliana*, 17 genes in *Hordeum vulgare*, 19 genes in *Oryza sativa*, 29 genes in *Populus trichocarpa*, and 31 genes in *Zea mays* (Cardon et al. 1997; Tripathi et al. 2018; Xie 2006; Jiao et al. 2010; Li and Lu 2014; Moreno et al. 1997; Wang et al. 2005). Besides these differences in the numbers of genes, *SPL* transcription factors also exhibit species-specific functions. For example, *O. sativa* *SPL8* (*OsSPL8*) controls the development of the auricle, which is a unique phenotype of monocotyledons (Moreno et al. 1997; Lee et al. 2007). Another study revealed that several *cis*-acting elements in the *Gossypium arboreum* and *H. vulgare* *SPL* promoter regions are associated with responses to light, stresses, and phytohormones (Cai et al. 2018; Tripathi et al. 2018). The numbers, structures, and functions of the *SPL* genes have also been studied in other major crops and in some model plant species, such as *Glycine max*, *Lycopersicon esculentum*, *Malus pumila*, *O. sativa*, *Petunia hybrida*, and *Salvia miltiorrhiza* (Tripathi et al. 2017; Manning et al. 2006; Miura et al. 2010; Li et al. 2013; Preston et al. 2016; Zhang et al. 2014). For example, in *O. sativa*, high expression of the *SPL* genes leads to grain yield enhancement and increases in crop production (Jiao et al. 2010; Miura et al. 2010).

Improvements in crop yields are urgently needed in economically important cultivated species of the Juglandaceae, such as *Carya cathayensis* (Chinese hickory), *Carya illinoensis* (pecan), and *Juglans regia* (Persian walnut). These nuts have known health benefits and can reduce the risk of cardiovascular disease and coronary heart disease (Hu et al. 1998; Wu et al. 2004; Wang et al. 2007). *C. illinoensis* is native to North America, and its excellent commercial value has led to worldwide planting on a large scale (Sparks 1992). *C. cathayensis* is the most important woody oil-bearing plant in southeastern China and provides important support to the local economy as well as widely recognized health benefits in the Chinese community. *J. regia* is the most widely consumed tree nut in the world. Thus, increasing the yield of these plants is of considerable worldwide interest.

In this work, we first mined the *SPL* gene family from the available genome, coding, and protein sequences in various

plant species, including members of the Juglandaceae. We then analyzed the evolution and function of the *SPL* gene family. Differences in the number of *SPL* genes among the Juglandaceae species were examined through the analyses of phylogenetic relationships, sequences and motifs, gene expression, and selection pressure. The data presented here may provide an important basis for future investigations of *SPL* genes, with implications for improving crop yield in the Juglandaceae.

Materials and methods

Identification of *SPL* gene family members

A total of 17 *A. thaliana* *SPL* genes and predicted proteins were retrieved from the TAIR database (<http://www.arabidopsis.org>) to serve as queries for the homologous search described below. The complete genome and predicted protein sequences of *Amborella trichopoda*, *A. thaliana*, *G. max*, *Medicago truncatula*, *O. sativa*, *Physcomitrella patens*, *P. trichocarpa*, *Vitis vinifera*, and *Z. mays* were obtained from the Phytozome database (version 12.0, <https://phytozome.jgi.doe.gov/pz/portal.html#>). The complete genome and predicted protein sequences of *J. regia* were obtained from the NCBI database (<https://www.ncbi.nlm.nih.gov>); the *J. regia* sequence name was modified with a combination of the Latin abbreviation and protein ID, such as JrXP_018805983.1. The genome draft and predicted protein sequences of *C. cathayensis* and *C. illinoensis* were obtained from our assembled data, which are currently being submitted. A search of *SPL* homologous sequences within these land plants was performed using local blastn, blastp (Altschul et al. 1990), and hmmsearch (Finn et al. 2011) with an *e* value less than 10^{-5} . These sequences (short sequences were discarded) were translated, and then domain searches against the Pfam database (Finn et al. 2010) were conducted using the SMART program (<http://smart.embl-heidelberg.de>). Candidate sequences lacking the SBP domain were excluded from phylogenetic analysis.

Phylogenetic analysis

All *SPL* amino acid sequences were aligned with the Muscle tool of MEGA 5.0 (Tamura et al. 2011). Default parameters were used to determine the best-fitting evolutionary models by the model test in MEGA 5.0. A phylogenetic relationship of the *SPL* protein tree was constructed according to the maximum-likelihood tree using RAXML-NG (<https://github.com/amkozlov/raxml-ng>) to obtain the number and composition of orthologous groups, and sequences with a significant mislocation were manually adjusted. The maximum-likelihood analysis was performed with the JTT+G

model with bootstrap values as 1000. The same family species of the same phylogenetic tree branch were joined into one group. The bootstrap level of each branch is not shown in the figure. However, the complete phylogenetic tree can be seen in supplementary figure S1. The motifs of the *SPL* genes were analyzed using MEME (<http://meme-suite.org/tools/meme>) with a maximum motif length of 100 and a search number of 20. The *SPL* genes of various species in each clade were also counted.

The reliability of the phylogenetic tree constructed by RAxML-NG was verified by also constructing phylogenetic trees based on *SPL* protein sequences using the neighbor-joining (NJ) and maximum-likelihood (ML) methods of MEGA5.0. The NJ and ML trees were constructed with the ‘pair-wise deletion’ option and the ‘number of bootstrap replications’ model, with a bootstrap test comprising 1000 replicates. The corresponding proteins from *A. trichopoda* (an ancient angiosperm) were used as an outgroup for the phylogenetic tree of Juglandaceae *SPL* genes, which was constructed with the NJ method.

Selection pressure analysis

A branch model was used to evaluate the ω -value of each clade with PAML (Yang 2007). A Likelihood Ratio Test (LRT) was performed to either accept or reject the hypothesis of model 0 (only one ω value in the tree) or model 2 (Two ratio Model 2: allow the ω value to vary among the foreground and background branch) among the different clades. The LRT was calculated assuming that twice the difference in the log of maximum-likelihood was distributed as a χ^2 distribution, with the degrees of freedom given by the difference in the number of parameters in the models. Each clade was set as the foreground for testing. The selection pressures of angiosperms were also calculated in the same way.

Positive selection was evaluated using a branch site model: Model A allowed positive selection on one or more branches and Model A null did not allow positive selection. Each clade was set as the foreground to test for positive selection sites with the LRT calculation. The Bayes Empirical Bayes analysis was used to evaluate the posterior codon probability of the positive selection site class on the foreground branch.

Transmembrane structural analysis

The possible functional structures at the C-terminal of the *SPL* proteins were searched using the SMART program with *C. cathayensis SPL* genes as representative genes. If a transmembrane structure was detected in the C-terminal, the transmembrane structure of all *SPL* proteins was then predicted with the TMHMM program (version 2.0, Krogh et al. 2001). For the predictions, the *SPL* proteins were considered

likely to contain a transmembrane region or signal peptide if the expected number of amino acids in the transmembrane helices (ExpAA) exceeded 18.

Expression analysis

The fragments per kilobase million (FPKM) values of *C. cathayensis* and *C. illinoensis* RNA sequencing data (submitted, Bioproject ID PRJNA435846) were utilized for the gene expression analysis of various tissues. The analyzed *C. cathayensis* tissues were male and female flower buds, male inflorescences, exocarps, endocarps, embryos, and vegetative shoots. Similarly, the examined *C. illinoensis* organs included exocarps, embryos, roots, stems, and leaves. The gene expression maps were modified from the analyzed results from the MultiExperiment Viewer (Howe et al. 2010).

Prediction of hickory *SPL* genes targeted by *miR156*

C. cathayensis miR156 (*miR156a*, *miR156b*, *miR156i*, *miR156j*, and *miR156k*) sequences were obtained from a recent study (Sun 2016). The *miR156* target site of *C. cathayensis SPL* genes was predicted by searching the 3' untranslated regions and coding regions of all *C. cathayensis SPL* genes using the psRNATarget with default parameters (Dai and Zhao 2011).

Results

Classification and evolution of the *SPL* family in angiosperms

In this study, we utilized 17 *A. thaliana SPL* (*AtSPL*) protein sequences as the queries for local blastp, blastn, and hmmsearch searches against our custom land plants database. Our results showed that the *SPL* gene family is common among land plants. All analyzed protein sequences contained a representative domain, which was identified as SBP by the SMART program.

A total of 293 *SPL* genes were identified through a comprehensive search from our customized database. The results indicated that *A. trichopoda* has the least number of *SPL* genes, while *J. regia* has 37 *SPL* genes (Table 1 and Table S1). The dicotyledonous model plant, *A. thaliana*, has a total of 17 *SPL* genes. However, we noticed that the *A. thaliana SPL AtSPL13A* gene differed from the *AtSPL13B* gene by only one base. Therefore, we assumed they are the same gene, so a total of 16 *A. thaliana SPL* genes were used in this study (Table 1). Our search results for *SPL* genes showed that fewer than 10 genes were detected by blastn and blastp in each Juglandaceae species. However, the

Table 1 Number and distribution of *SPL* genes in the Juglandaceae and other land plants

Family	Species	Gene ID	Number											total
			Clade1.1	Clade1.2	Clade1.3	Clade2	Clade3.1	Clade3.2	Clade3.3	Clade3.4	Clade3.5	Clade3.6		
Funariaceae	<i>Physcomitrella patens</i>	Phpat	1	2	0	0	5	0	3	0	0	0	11	
Amborellaceae	<i>Amborella trichopoda</i>	scaffold	1	1	1	1	1	1	0	1	1	1	9	
Gramineae	<i>Zea mays</i>	GRMZM	1	2	4	0	5	1	5	4	4	0	26	
	<i>Oryza sativa</i>	LOC.Os	1	1	2	0	3	1	4	1	4	0	17	
Brassicaceae	<i>Arabidopsis thaliana</i>	At	1	2	2	3	1	0	3	2	1	1	16	
	<i>Arabidopsis lyrata</i>	Aly	1	2	2	2	1	0	3	2	1	1	15	
Vitaceae	<i>Vitis vinifera</i>	VIT	1	1	2	3	1	1	2	1	3	2	17	
Salicaceae	<i>Populus trichocarpa</i>	Potri	2	2	4	4	2	2	3	1	4	3	27	
Leguminosae	<i>Glycine max</i>	Glyma	2	2	6	7	2	4	2	3	6	4	38	
	<i>Medicago truncatula</i>	Medtr	1	1	3	4	1	1	4	2	1	2	20	
	<i>Carya cathayensis</i>	Cc	1	2	4	6	2	2	2	2	5	4	30	
Juglandaneae	<i>Carya illinoensis</i>	CIL	2	2	4	5	2	1	2	2	6	4	30	
	<i>Juglans regia</i>	Jr	3	1	5	5	2	2	4	2	10	3	37	
total			18	21	39	40	28	16	37	23	46	25	293	

hmmsearch results for the SBP domain revealed more than 30 genes in each Juglandaceae species. Our results suggested that the *SPL* gene sequences in Juglandaceae species differ considerably from those of *A. thaliana*. Certain family members with diverse sequences could easily be missed if only counting using a sequence similarities search.

We investigated the phylogenetic relationships among the *SPL* genes of the aforementioned species using 293 *SPL* genes to construct phylogenetic trees and subsequently analyzed the motifs (Fig. 1 and Fig. S1). The 293 *SPL* genes were divided into three major clades according to sequence lengths, as well as the combination of motifs and topological structures of phylogenetic trees. Clade 1 *SPL* genes encoded long *SPL* proteins, while the *SPL* genes of Clades 2 and 3 encoded short *SPL* proteins. Clades 1 and 3 contained multiple motifs that could be further divided into three and six smaller branches, i.e., Clades 1.1–1.3 and Clades 3.1–3.6, respectively. Although the *SPL* gene family has expanded in land plants, our analyses of the encoded protein sequences indicated that this gene family follows a conserved evolutionary model. This observation is in accord with a previous study by Preston and Hileman (2013). We obtained an evolutionary relationship point of view by constructing the phylogenetic trees using the NJ method (Fig. S2) and the ML method (Fig. S3). We noticed that the position of the six branches in Clade 3 differed somewhat depending on the method used, but the evolutionary relationships in each clade were conserved.

Next, we carried out a motif analysis to assess the reliability of the obtained phylogenetic grouping. We analyzed the composition and distribution of motif-encoding sequences in the *SPL* genes, and our results showed that the motif of the *SPL* genes was highly specific among the different clades (Fig. 1). The MEME results indicated that all *SPL* genes contained motif 1. The *SPL* genes in Clade 2 contained only motif 1. By contrast, Clade 1 *SPL* genes

contained motifs 1 to 3, while all Clade 3 *SPL* genes contained motifs 1 and 5. Analysis of the motif sequences revealed that motif 1 contained the sequences typical of the SBP domain (i.e., encoding Zn1, Zn2, and NLS). Thus, we concluded that motif 1 is the SBP domain. The *SPL* motifs were highly specific among the different clades. However, the motif composition in the N-terminal was relatively highly conserved.

Although the *SPL* gene family has expanded in angiosperms, analyses of the encoded protein sequences indicated that this gene family followed a conserved evolutionary model. The motif analysis results were consistent with the grouping results of the topology of the phylogenetic tree. For some *SPL* genes with an incomplete SBP domain sequence, we speculated that the encoded proteins consisted of a partial SBP structure. Because the missing part was not conserved, these sequences were not detected as a motif by MEME. The specific cause of this phenomenon requires further study. However, all these sequences contained the SBP domain sequences according to a SMART search, implying that they are *SPL* genes. Based on the MEME results, we explored the *SPL* gene distribution in greater detail in the plant species mentioned above in each clade. We observed that Clade 2 and Clade 3.6 did not contain *SPL* genes derive from Gramineae species, whereas Clade 3.2 lacked the *SPL* genes from Brassicaceae species. The *SPL* genes of *P. patens* were observed in only a few clades. This distribution may reflect the deletion of genes during evolution, with the ancient species affected the most. All phylogenetic trees constructed with RAXML-NG (Fig. 1) as well as the NJ (Fig. S2) and ML (Fig. S3) methods indicated that Clade 2 was highly conserved, and lacked the monocotyledon *SPL* genes, whereas sequences from the other species were well preserved.

Fig. 1 Phylogenetic relationship and motif patterns of the *SPL* gene family in the Juglandaceae and other land plants. The phylogenetic tree was built using the JTT+G model with 293 *SPL* protein sequences. The *SPL* genes of the same family were combined into one aggregate. Conserved motifs among *SPL* proteins were identified through MEME analysis. The SBP domain (motif 1) is marked with a gray box, while the other motifs are marked with black boxes

Evolutionary selection of *SPL* genes in Juglandaceae

The number of *SPL* genes in Juglandaceae species was similar to that of *G. max*, *P. trichocarpa*, and *Z. mays* and was considerably higher than in most other land plants examined (Table 1). The *SPL* genes of *G. max*, *P. trichocarpa*, and *Z. mays* were reported to have undergone a specific duplication event (Ralph et al. 2008; Schmutz et al. 2010; Hirsch et al. 2016). Therefore, we also constructed a phylogenetic tree of the Juglandaceae *SPL* genes to investigate whether these genes had undergone a Juglandaceae-specific duplication event within specific species. Our results indicated that only six pairs of *J. regia* *SPL* genes existed in the form of paralogous gene pairs, corresponding to 16.22% of the total number of *J. regia* *SPL* genes, but they were not present in *C. cathayensis* and *C. illinoensis* (Fig. S1). Conversely, 94.74% of the

G. max and 74.07% of the *P. trichocarpa* *SPL* genes were detected as paralogous gene pairs (Fig. S1). These observations suggested that the Juglandaceae *SPL* genes had not undergone a Juglandaceae-specific duplication within species. Their considerable abundance is likely to reflect the occurrence of one replication before differentiation in the common ancestors.

Many paralogous *SPL* genes generated by replication events may have been affected during evolution by functional changes, including neofunctionalizations and subfunctionalizations. We speculated that *SPL* genes in some clades underwent an adaptive evolution, with corresponding changes to the encoded amino acid sequences. We tested this hypothesis by investigating selection models of all clades using Codeml (Yang 2007). Branching models were used to investigate the selection patterns of the small branches in Clades 1–3.

Fig. 2 Phylogenetic tree of Juglandaceae *SPLs* and branch selection pressure. Phylogenetic trees were constructed by the NJ method using nucleic acid sequences. **a** Selection pressure of Clade 1. **b** Selection pressure of Clade 2. **c** Selection pressure of Clade 3. The ω -value is

Our results showed that, in general, the ω -value of all branches was less than one (Fig. 2), indicating that all *SPL* genes were under purifying selection pressure and that they had evolved in a relatively conservative manner. Generally, Clade 2 and most of the branches of Clades 1 and 3 had lower ω -values, except for Clade 1.3 and Clade 3.5 (Fig. 2). The ω -value differed substantially within the branches, even in the same primary clade. For example, the ω -value of the highly conserved Clade 2 varied up to 10 fold among the different branches. Early differentiated branches generally had lower ω -values, except for Clade 2. With the constant differentiation, the ω -value of Clade 2 branches reduced (Fig. 2b). In Clade 1, the earlier emerging branches were more conserved. The replication and subsequent functional differentiation of the genes in new branches were due to plant adaptations to the environment (Fig. 2a). Clade 3, which included genes encoding the short *SPL* proteins, exhibited similar evolutionary characteristics, with ω -values between 0.100 and 0.150, except for clade 3.5 (Fig. 2c).

Some species had experienced special natural and differing selection pressures. We tested whether a substantial difference existed between the Juglandaceae and other species by analyzing the branching models between angiosperms and *J. regia* plants (Fig. S4). The ω -values were higher for the Juglandaceae than for the angiosperms, especially in Clades 1.1–1.3 and Clade 3.5. The ω -values of the

the selection pressure of the branch: $\omega < 1$ the branch is subjected to purification selection, $\omega = 1$ the branch is subjected to neutral selection, and $\omega > 1$ is the branch is subjected to evolutionary selection

four branches of the Juglandaceae were more than 10 times greater than the ω -values of the angiosperms (Fig. 2a, c, S4).

Juglandaceae *SPL* gene selection sites

We performed Bayes Empirical Bayes analysis (Yang et al. 2005) to predict the selection sites of the Juglandaceae *SPL* genes in Clades 1 and 3, with each clade as a foreground and all clades in same major groups as the background. We detected 17 selection sites in the Clade 1 genes (Table S2) and seven selection sites in the Clade 3 genes (Table S3). These sites had a posterior probability greater than 0.9 under positive selection pressure (ω -value > 1). All selection sites of the Clade 3 genes were located in the SBP domain sequence (Fig. 3), with 23 K (Clade 3.5) located in Zn1, 61D (Clade 3.5) and 62G (Clade 3.1) located in Zn2, and 75E (Clades 3.4 and 3.6) located in the NLS. Changes in Zn1, Zn2, and the NLS affect the ability of the *SPL* proteins to bind to DNA, thereby leading to changes in gene function and differentiation of these branches.

Most of the Clade 1 selection sites were located at the C-terminal. The Clade 1.2 genes had the most selection sites, but their SBP domain at the N-terminal contained only four selection sites (7R, 9S, 45S, and 67S; Table S2). Although the sequence corresponding to the N-terminal was conserved, many selection sites had been generated.

Fig. 3 Selective pressure in specific sites of Clade 3. The proteins in this clade that have been subjected to a highly significant selection pressure are highlighted with red stars. The proteins subjected to a significant selection pressure are marked with yellow stars

The C-terminal was even less conserved, so selection sites were likely to be generated more easily. The C-terminal of the Clade 1 genes carried multiple selection sites, but no reports have described an *SPL* gene with domains other than the SBP domain. A motif analysis by MEME revealed that the C-terminal of the Clade 1 genes has multiple conserved motifs (motifs 2, 3, 4, 8, 10, and 12), which are unique to this branch (Fig. 1). The Pfam search for the Clade 1 domain indicated that this motif's combinations at the C-terminal did not match any known domain.

Interestingly, we detected a transmembrane structure at the C-terminal of the *SPL* proteins. Hence, we used the TMHMM program to predict the transmembrane structure encoded by all *SPL* genes. The protein is considered to contain a transmembrane region if the ExpAA value is greater than 18. We observed that 85% of the *SPL* genes in Clade 1 encoded a transmembrane structure at the C-terminus, accounting for 77% of all the *SPL* genes that have a transmembrane structure (Table S1). This structure is likely to be specific to the proteins corresponding to the Clade 1 genes. Further study is needed to verify this other potential functional structure in the conserved motif combinations.

Expression analysis of Juglandaceae *SPL* genes

We conducted an expression analysis of the *C. cathayensis* and *C. illinoensis* *SPL* genes to study the possible functional roles of these genes (Fig. 4). The *C. cathayensis* *SPL* genes were highly expressed throughout the differentiation stages of the male and female flower buds (Fig. 4a), suggesting that *SPL* genes were involved in *C. cathayensis* flower development. The *SPL* genes were also expressed

in vegetative shoots and leaves, exocarp, endocarp, and embryo (Fig. 4). We noticed a variation in the expression level of *C. cathayensis* *SPL* genes in different branches of the same clade (e.g., Clades 2, 3.2, and 3.5; Fig. 4a). The gene expression level of one branch was either decreased considerably or the gene was no longer expressed. The *SPL* genes of these clades were not expressed in other tissues or growth periods. The *C. illinoensis* *SPL* expression levels also exhibited similar profiles (e.g., Clade 3.3 and 3.5; Fig. 4b). These genes with decreased expression levels may have been affected by defunctionalizations or subfunctionalizations during evolution.

Among the three clades, the Clade 1 genes of *C. cathayensis* and *C. illinoensis* exhibited a similar expression pattern to that of *L. esculentum* *SBP7* (Salinas et al. 2012). These genes were widely expressed in flower buds, pericarps, embryos, leaves, vegetative shoots, and roots. The motif combinations of the N-terminal of the three clades were highly conserved, so they were unlikely to cause the observed differences in expression of the *SPL* genes. However, the branch-specific motif combinations at the C-terminus of Clade 1 (Fig. 1) was most likely to cause this type of change. Therefore, we have proposed that the expression of Clade 1 genes was due to the C-terminal motif combinations. Evidence from the present study also suggested that the C-terminal motif combinations and multiple selection sites located in the motif combinations might contribute to the expression of Clade 1 genes in various plant tissues.

Interestingly, Clade 3.5, which has the highest ω -value (Fig. 2c), differed from the other Clade 3 branches in *C. illinoensis*. Some of the *SPL* genes in Clade 3.5 have higher expression levels in roots (Fig. 4b). No similar profile was

Fig. 4 Expression profiles of *Carya cathayensis* and *Carya illinoensis* *SPL* genes. **a** Expression patterns of *C. cathayensis* *SPL* genes. **b** Expression patterns of *C. illinoensis* *SPL* genes. *H* stands for *C. cathayensis*, *P* stands for *C. illinoensis*. The middle four digits indicate the sampling data. March 11th, March 18th, and April 8th are the three stages before, during, and after differentiation of *C. cathayensis*

observed in the *C. cathayensis* expression data. Between the different branches of the same major group, the expression pattern of the *C. cathayensis* *SPL* gene is generally consistent (Fig. 4b).

Discussion

The *SPL* gene family plays an essential role in the regulation of plant growth and flowering (Wang et al. 2009), and it exhibits species-specific functions. For example, *OsSPL8* can control the development of auricles, which are unique structures to the Gramineae (Moreno et al. 1997; Lee et al. 2007). In the Juglandaceae, the male flowers of *C. cathayensis*, *C. illinoensis*, and *J. regia* become dormant after differentiation, and blossoming occurs in the next spring (Latorse 1985; Zhang et al. 2008; Xie 2013). By contrast, the *SPL* genes are induced throughout the development of the entire male flower. Regardless of whether the *SPL* genes regulate this particular developmental process in the Juglandaceae, understanding the role of *SPL* genes in the Juglandaceae in functional or structural differentiation is of considerable research interest. However, our understanding of the role of the *SPL* genes is still limited. Therefore, the first step in understanding the functional diversity of these genes is to study the evolutionary history of the *SPL* gene family.

sis female flower buds, and May 8th to June 13th represents the male bud differentiation period for *C. cathayensis*. FPKM fragments per kilobase million, *EM* embryo, *EN* endocarp, *EX* exocarp, *FF* female flower bud, *L* leaf, *MF* male flower bud, *MI* male inflorescence, *R* root, *VB* vegetative branch

Evolution and expansion of the Juglandaceae *SPL* gene family

Our results indicate that the *SPL* gene family has stabilized and is moving in an increasingly conservative direction (i.e., the ω -values are less than 1; Fig. 2 and Fig. S4). However, the number of *SPL* genes varies from species to species. Phylogenetic evidence suggests that the gene encoding the SBP domain has undergone several ancient as well as more recent replication events, leading to the formation and retention of multiple *SPL* paralogs and branches (Preston and Hileman 2013). Ancient plant species contain relatively few *SPL* genes, suggesting that multiple gene duplications have occurred during the evolution of the *SPL* gene family (Kaufmann et al. 2010). The *SPL* genes of *G. max*, *P. trichocarpa*, and *Z. mays* underwent specific duplications, resulting in significantly more *SPL* genes in their genomes than are found in other species in the same family (Ralph et al. 2008; Schmutz et al. 2010; Hirsch et al. 2016). Although the Juglandaceae *SPL* genes did not undergo species-specific duplication, the duplication experienced by their ancestors prior to differentiation may have resulted in a similar differentiation potential and number of members as species that did undergo species-specific duplications (Table 1 and Fig. S1). Whether functional changes accompany these gene duplication events

remains to be determined. Future studies should look for new differences in these gene sequences and structures.

We observed that the genetic relationship of the *SPL* genes between different species might not match the genetic relationship of other species, such as Clade 2 (Fig. 1 and Fig. S1). For example, *A. trichopoda* is a relict plant that separated from other angiosperms before it differentiated. In Clade 2, the diversification of some *A. trichopoda* *SPL* genes occurs after that of the Juglandaceae. This phenomenon may reflect a different rate of evolution of the *SPL* genes from different species under specific environmental conditions. When the evolutionary rate is highly inconsistent, rapidly evolving species may introduce multiple mutations at certain sites, so the bases remain the same. These sites should appear relatively late in the evolution process. When this non-homologous similarity was sufficient to mask the true historical information in the sequence, it created a conflict between the phylogenetic trees based on gene or protein sequences and the relationships among the evolved species (Felsenstein 1978; Andreassen and Baldwin 2001). Conversely, it also reflected the evolution rate of the earliest branches in Clade 2, which has three Juglandaceae *SPL* genes and one *V. vinifera* *SPL* gene that differed from the other *SPL* genes on the same clade. We observed that the *SPL* genes of *C. cathayensis*, *C. illinoensis*, and *J. regia* lacked an apparent differentiation order during evolution. Among these three species, *C. cathayensis* and *C. illinoensis* showed the closest phylogenetic relationship. However, in some branches, such as Clades 3.2, 3.3, and 3.4 (Fig. 2), the *J. regia* *SPL* genes were closely related to one of them.

While constructing the phylogenetic tree of the Juglandaceae Clade 3 for pressure calculations, we found that Clade 3.6 (which should have been grouped with Clade 3.5) was clustered with Clades 3.3 and 3.4 (Fig. 2c). A phylogenetic relationship such as this may be due to a deletion. We need to delete the non-shared motif structure, leaving only the members that have motifs 1 and 5 (Fig. 1) prior to the branch model calculation. Our results indicated that the motifs 1 and 5 sequences of the Clade 3.6 *SPL* genes were most similar to the corresponding motif sequences in Clades 3.3 and 3.4. The Clades 3.5 and 3.6 are clustered together in a complete phylogenetic tree (Fig. 1) most likely because of the high similarity of parts other than motifs 1 and 5.

Differentiation of the *SPL* gene family

Gene duplication is a main driving force for the acquisition of new genes and the creation of new functions (neofunctionalizations and subfunctionalizations) (Taylor and Raes 2004). A central theory of molecular evolution states that most genes evolved primarily under strong purifying constraints to conserve functions. Gene duplication allows a gene to be free from this selection pressure and to eventually

accumulate mutations that can lead to new functions or a complete loss of function (Prince and Pickett 2002).

The *SPL* genes of the Juglandaceae species had all been subjected to purifying selection pressure, but the variation in the ω -values among different branches suggests that the evolution of the genes in these branches was not synchronized. Genes with lower ω -values encode important proteins that may be beneficial to plants and may be under purifying selection pressure, while genes with higher ω -values are under strong selection pressure and may have evolved new functions or become pseudogenes. Selection pressure differences are also likely to be responsible for differences in the gene evolution rate of each branch.

Our results showed that the Clade 2 and 3 *SPL* expression patterns (Fig. 4) were similar to those of many other species, with relatively high expression levels in inflorescences and flower buds (Xie 2006; Salinas et al. 2012; Li et al. 2013). By contrast, Clade 1 *SPL* genes (encoding lone *SPL* proteins) were highly expressed in all tissues. This expression pattern differed from that of other clades, indicating that functional differentiation has taken place. Previous studies have shown that these genes are essential for plant growth and development and cannot be easily changed. For example, the *A. thaliana* *SPL* genes (*AtSPL7*, *AtSPL14*, and *AtSPL16*) in this clade are important for certain activities, including regulation of copper homeostasis in the root (Garciamolina et al. 2014a, b), plant development, and sensitivity to fumonisin B1 (Stone et al. 2010), respectively. Genes in this clade usually did not undergo functional differentiation unless exposed to major changes in the external environment. The three branches of Clade 1 have been widely studied in terms of the functional annotation and subsequent validation of members in Clades 1.1 and 1.2, while relatively few studies have been conducted on members in clade 1.3 (Preston and Hileman 2013). Our results showed that Clade 1.3 has the most members and the highest ω -values in Clade 1, suggesting that Clade 1.3 has the most diverse potential in Clade 1. The expression patterns of Clade 1.3 did not differ much from those of the other two branches, so the Clade 1.3 genes may redundantly regulate plant growth and development along with the genes in the other branches in Clade 1. Lan et al. (2017) described that an individualized genomic architecture might play an important role in adaptation to specific environments when plants are under purifying selection. From the above observation, we speculate that when an environment changes considerably, the *SPL* genes of Clades 1.1 and 1.2 will continue to play their original roles, while those of Clade 1.3 can develop further. Therefore, Clade 1.3 is most likely a newly differentiated branch of Clade 1 in the evolutionary process.

The selection site loci in the Clade 1 genes existed in the sequences encoding the SBP domain and the C-terminal conserved motif combination. The presence of Zn1, Zn2,

and NLS in the SBP domain implies that this domain is important for DNA binding (Yamasaki et al. 2004; Birkenhil et al. 2005). Single amino acid substitutions in the SBP domain, and especially in Zn1, Zn2, and NLS, can alter the ability of SPL to bind to DNA. However, the temporal and spatial *SPL* expression levels are not influenced by the SBP domain sequence. Therefore, any observed temporal and spatial differences in expression are most likely due to the branch-specific motif combination encoded in the sequence corresponding to the C-terminus. We detected a transmembrane structure in the conserved motif combination at the C-terminus of Clade 1 *SPL* proteins. Interestingly, no Pfam domain matches this cascade. These observations do not mean that no functional structure exists in this region; therefore, further studies should focus either on mutating the selection site in the region or deleting the region to explore its actual function.

Clade 3 has the largest number of members among the three major clades, each with its own motif structure and functionality. In *C. cathayensis*, the *SPL* genes were highly expressed throughout the development of the male and female flowers, suggesting that these genes are closely related to flower development (Fig. 4a). In *C. illinoensis*, Clade 3.5 has the highest ω -value in the *SPL* gene family, while the other five branches have similar ω -values between 0.1 and 0.2 (Fig. 2). Expression analysis revealed that the branches of Clade 3 have similar expression patterns, except for Clade 3.5 (Fig. 4), and the *SPL* genes in Clade 3.5 were highly expressed in the roots (Fig. 4b). The *O. sativa* *SPL* genes in Clade 3.5 (Fig. S1) can alter the pattern of cell proliferation and elongation, leading to changes in grain shape (Wang et al. 2012). The *Z. mays* *SPL* genes in Clade 3.5 (Fig. S1) show a greater expression in mutant feminized tassels than in wild type, suggesting a role in ear-specific traits (Hultquist and Dorweiler 2008). The function of the *SPL* genes in Clade 3.5 is species specific and displays large differences in tissue specificity in *C. illinoensis*. Determining whether similar functions could be assigned to the Juglandaceae *SPL* genes would, therefore, be of great interest. Generally, when compared with angiosperms, the purifying selection pressure of the Juglandaceae *SPL* genes is small, indicating that the Juglandaceae *SPL* genes are more susceptible to functional differentiation.

Expression analysis also indicated differences in the expression levels of *SPL* genes within the same clade (e.g., Clades 2 and 3). These genes showed either decreased expression or were no longer expressed. These *SPL* genes were also not expressed in other tissues and periods (Fig. 4). The reduced expression levels may be affected by defunctionalizations and subfunctionalizations during evolution. One possible hypothesis is that the moderate expression level of the *SPL* genes (assumed to be homologous) is due to the redundancy of their functions. These genes are maintained in

the plant genome and can compensate for mutations that may occur during evolution, so that plant growth is unaffected and the mutations have no obvious side effects.

***SPL* genes are regulated by miR156**

Previous research has revealed that the genes encoding short *SPL* proteins are largely regulated by *miR156* and *miR157*, with their complementary sites at the 3' untranslated or coding regions of the genes (Matthew et al. 2002). Our results suggested that 88% of *C. cathayensis* *SPL* genes (15 genes) in Clade 3, except for Clade 3.1, was regulated by *miR156* (Fig. S5 and Table S6). Only a few of the *C. cathayensis* *SPL* genes were clustered into Clade 3.1 (Fig. S1), which contained *SPL* genes of *A. thaliana* (*AtSPL8*) and *P. trichocarpa* (*PtSPL21* and *PtSPL26*). Previous studies have concluded that *AtSPL8*, *PtSPL21*, and *PtSPL26* were not regulated by either *miR156* or *miR157* (Preston and Hileman 2013; Li and Lu 2014). Therefore, the lack of regulation of Clade 3.1 *SPL* genes by *miR156* appears to be universal among species. The *miR156* target sequence is located at base positions 681 to 2081 (encoding amino acids 227 to 693), and none of Clade 3 *SPL* genes have a *miR156*-targeted site in their SBP domain (Table S6). This observation suggests that the *miR156* regulatory site is not affected by selection sites. The *miR156* regulatory sites are not easily affected even when species have evolved under environmental selection pressures.

Conclusions

In this study, we performed a phylogenetic analysis of the *SPL* gene family of the Juglandaceae and other land plant species. We identified 293 *SPL* genes in 13 species, with *J. regia*, *C. cathayensis*, and *C. illinoensis* containing 30, 30, and 37 *SPL* genes, respectively. We have classified the *SPL* gene family into three major clades, and further divided them into 10 small clades. All clades of the *SPL* gene family have been purified and have evolved to stabilize. However, the purifying selection of the Juglandaceae *SPL* genes was smaller than that of the angiosperms. Our results indicated that the Juglandaceae *SPL* gene is more susceptible to functional differentiation than are the *SPL* genes of most angiosperms. The Clade 3.5 has the highest ω -value and the greatest differentiation potential. Expression analysis indicated that the *SPL* genes of Clade 3.5 were indeed different from the *SPL* genes of other branches in Clade 3 and are highly expressed in roots. It will be interesting to see if similar functions can be assigned to the Juglandaceae *SPL* genes in this clade. The expression analysis also showed that Clade 1 and the other two branches show spatial and temporal differences. Clade 1 genes are highly expressed in various tissues

and organs, and this expression profile is closely related to many motif combinations at the C-terminal. We also observed a branch-specific membrane structure and many selection sites at the C-terminal of Clade 1. This structure and these sites will be important entry points for revealing the functional differentiation of the *SPL* gene family.

Acknowledgements We thank Dr. Qingpo Liu of Zhejiang A&F University, Dr. Fei Chen of Fujian Agriculture and Forestry University for their collaboration and support.

Author contributions ZJW and LSZ conceived and designed the research. DW designed, planned and carried out the experiments. DW, JL, and JHC performed data analysis. DW contributed in data interpretation. KJL and DW wrote and revised the manuscript. All authors read, commented and approved.

Funding This study was funded by Zhejiang Provincial Natural Science Foundation of China (LY15C160003) and National Natural Science Foundation of China (31600547).

Compliance with ethical standards

Conflict of interest All the authors declare that there is no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Andreasen K, Baldwin BG (2001) Unequal evolutionary rates between annual and perennial lineages of checker mallows (*Sidalcea*, Malvaceae): evidence from 18S-26S rDNA internal and external transcribed spacers. *Mol Biol Evol* 18:936–944
- Birkenbihl RP, Jach G, Saedler H, Huijser P (2005) Functional dissection of the plant-specific SBP-domain: overlap of the DNA-binding and nuclear localization domains. *J Mol Biol* 352:585–596
- Cai C, Guo W, Zhang B (2018) Genome-wide identification and characterization of SPL transcription factor family and their evolution and expression profiling analysis in cotton. *Sci Rep* 8:762
- Cardon GH, Höhmann S, Nettesheim K, Saedler H, Huijser P (1997) Functional analysis of the arabidopsis thaliana SBP-box gene SPL3: a novel gene involved in the floral transition. *Plant J* 12:367–377
- Cardon G, Höhmann S, Klein J, Nettesheim K, Saedler H, Huijser P (1999) Molecular characterisation of the arabidopsis SBP-box genes. *Gene* 237:91–104
- Dai X, Zhao PX (2011) PsRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res* 39:W155–W159
- Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. *Syst Zool* 27:401–410
- Finn RD, Mistry J, Tate J, Coghill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer EL, Eddy SR, Bateamn A (2010) The pfam protein families database. *Nucleic Acids Res* 32:D211–D222
- Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 39:W29–W37
- Garciamolina A, Xing S, Huijser P (2014a) A conserved KIN17 curved DNA-binding domain protein assemblies with SQUAMOSA PROMOTER-BINDING PROTEIN-LIKE7 to adapt *Arabidopsis* growth and development to limiting copper availability. *Plant Physiol* 164:828–840
- Garciamolina A, Xing S, Huijser P (2014b) Functional characterisation of *Arabidopsis* SPL7 conserved protein domains suggests novel regulatory mechanisms in the Cu deficiency response. *BMC Plant Biol* 14:231
- Gou JY, Felippes FF, Liu CJ, Weigel D, Wang JW (2011) Negative regulation of anthocyanin biosynthesis in *Arabidopsis* by a miR156-targeted SPL transcription factor. *Plant Cell* 23:1512–1522
- Graaff EVD, Laux T, Rensing SA (2009) The WUS homeobox-containing (WOX) protein family. *Genome Biol* 10:248
- Guo AY, Zhu QH, Gu X, Ge S, Yang J, Luo J (2008) Genome-wide identification and evolutionary analysis of the plant specific SBP-box transcription factor family. *Gene* 418:1–8
- Hirsch CN, Hirsch CD, Brohammer AB, Bowman MJ, Soifer I, Barad O, Shem-Tov D, Baruch K, Lu F, Hernandez AG, Fields CJ, Wright CL, Koehler K, Springer NM, Buckler E, Buell CR, Leon BD, Kaeppler SM, Childs KL, Mikelg MA (2016) Draft assembly of elite inbred line PH207 provides insights into genomic and transcriptome diversity in maize. *Plant Cell* 28:2700–2714
- Howe E, Holton K, Nair S, Schlauch D, Sinha R, Quackenbush J (2010) MeV: MultiExperiment Viewer. In: Ochs M, Casagrande J, Davuluri R (eds) Biomedical informatics for cancer research. Springer, Boston
- Hu FB, Stampfer MJ, Manson JE, Rimm EB, Colditz GA, Rosner BA, Speizer FE, Hennekens CH, Willett WC (1998) Frequent nut consumption and risk of coronary heart disease in women: prospective cohort study. *BMJ* 317:1341–1345
- Hultquist JF, Dorweiler JE (2008) Feminized tassels of maize mopl1 and ts1 mutants exhibit altered levels of miR156 and specific SBP-box genes. *Planta* 229:99–113
- Jiao Y, Wang Y, Xue D, Wang J, Yan M, Liu G, Dong G, Zeng D, Lu Z, Zhu X, Qian Q, Li J (2010) Regulation of OsSPL14 by OsMiR156 defines ideal plant architecture in rice. *Nat Genet* 42:541–544
- Kaufmann K, Wellmer F, Muiño JM, Ferrier T, Wuest SE, Kumar V, Serrano-Mislata A, Madueño F, Krajewski P, Meyerowitz EM, Angenent GC, Riechmann JL (2010) Orchestration of floral initiation by APETALA1. *Science* 328:85–89
- Krogh A, Larsson B, Heijine GV, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J Mol Biol* 305:567–580
- Lan T, Renner T, Ibarra-Laclette E, Farr KM, Chang TH, Cervantes-Pérez SA, Zheng C, Sankoff D, Tang H, Purbojati RW, Putra A, DrautzMoses DI, Schuster SC, HerreraEstrella L, Albert VA (2017) Long-read sequencing uncovers the adaptive topography of a carnivorous plant genome. *Proc Natl Acad Sci USA* 114:E4435–E4441
- Latorse MP (1985) Etude de divers aspects de la reproduction sexuée chez la Noyer (*Juglans regia*. L) Dissertation, Université de Bordeaux II
- Lee J, Park JJ, Kim SL, Yim J, An G (2007) Mutations in the rice liguleless gene result in a complete loss of the auricle, ligule, and laminar joint. *Plant Mol Biol* 65:487–499

- Li C, Lu S (2014) Molecular characterization of the SPL gene family in *Populus trichocarpa*. *BMC Plant Biol* 14:131
- Li J, Hou H, Li X, Xiang J, Yin XJ, Gao H, Zheng Y, Bassett CL, Wang XP (2013) Genome-wide identification and analysis of the SBP-box family genes in apple (*Malus × domestica* Borkh.). *Plant Physiol Biochem* 70:100–114
- Manning K, Tör M, Poole M, Hong Y, Thompson AJ, King GJ, Giovannoni JJ, Seymour GB (2006) A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat Genet* 38:948–952
- Matthew WR, Brenda JR, Lee PL, Christopher BB, Bonnie B, David PB (2002) Prediction of plant microRNA targets. *Cell* 110:513–520
- Miura K, Ikeda M, Matsubara A, Song XJ, Ito M, Asano K, Matsukawa M, Kitano H, Ashikari M (2010) OsSPL14 promotes panicle branching and higher grain productivity in rice. *Nat Genet* 42:545–549
- Moreno MA, Harper LC, Krueger RW, Dellaporta SL, Freeling M (1997) *Liguleless1* encodes a nuclear-localized protein required for induction of ligules and auricles during maize leaf organogenesis. *Genes Dev* 11:616–628
- Preston JC, Hileman LC (2013) Functional evolution in the plant SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL) gene family. *Front Plant Sci* 4:1–13
- Preston JC, Jorgensen SA, Orozco R, Hileman LC (2016) Paralogous *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* (SPL) genes differentially regulate leaf initiation and reproductive phase change in petunia. *Planta* 243:429–440
- Prince VE, Pickett FB (2002) Splitting pairs: the diverging fates of duplicated genes. *Nat Rev Genet* 3:827–837
- Ralph SG, Chun HJ, Cooper D, Kirkpatrick R, Kolosova N, Gunter L, Tuskan GA, Douglas CJ, Holt RA, Jones SJ, Marra MA, Bohlmann J (2008) Analysis of 4,664 high-quality sequence-finished poplar full-length cDNA clones and their utility for the discovery of genes responding to insect feeding. *BMC Genom* 9:57
- Riechmann JL, Meyerowitz EM (1998) The AP2/EREBP family of plant transcription factors. *Biol Chem* 379:633–646
- Salinas M, Xing S, Höhmann S, Berndtgen R, Huijser P (2012) Genomic organization, phylogenetic comparison and differential expression of the SBP-box family of transcription factors in tomato. *Planta* 235:1171–1184
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinzaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463:178–183
- Sparks D (1992) Pecan cultivars: the orchard's foundation. *Pecan Production Innovations*, Watkinsville, pp 20–21
- Stone JM, Liang X, Neel ER, Stiers JJ (2010) Arabidopsis AtSPL14, a plant-specific SBP-domain transcription factor, participates in plant development and sensitivity to fumonisin B1. *Plant J Cell Mol Biol* 41:744–754
- Sun ZC (2016) The research of miRNA function in hickory female development. Dissertation, Zhejiang A&F University
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731–2739
- Taylor JS, Raes J (2004) Duplication and divergence: the evolution of new genes and old ideas. *Annu Rev Genet* 38:615–643
- Tripathi RK, Goal R, Kumari S, Dahuja A (2017) Genomic organization, phylogenetic comparison, and expression profiles of the SPL family genes and their regulation in soybean. *Dev Genes Evol* 227:1–19
- Tripathi RK, Bregitzer P, Singh J (2018) Genome-wide analysis of the SPL/miR156 module and its interaction with the AP2/miR172 unit in barley. *Sci Rep* 8:7085
- Wang H, Nussbaum-Wagler T, Li B, Zhao Q, Vigouroux Y, Faller M, Bombles K, Lukens L, Doebley JF (2005) The origin of the naked grains of maize. *Nature* 436:714–719
- Wang ZY, Liu R, Zhao X (2007) Plant resources. China Science and Technology Press, Beijing
- Wang JW, Czech B, Weigel D (2009) Mir156-regulated spl transcription factors define an endogenous flowering pathway in arabidopsis thaliana. *Cell* 138:738–749
- Wang SK, Wu K, Yuan QB, Liu XY, Liu ZB, Lin XY, Zeng RZ, Zhu HT, Dong GJ, Qian Q, Zhang GQ, Fu XD (2012) Control of grain size, shape and quality by OsSPL16 in rice. *Nat Genet* 44:950–954
- Wu X, Beecher GR, Holden JM, Haytowitz DB, Gebhardt SE, Prior RL (2004) Lipophilic and hydrophilic antioxidant capacities of common foods in the united states. *J Agric Food Chem* 52:4026–4037
- Xie K (2006) Genomic organization, differential expression, and interaction of SQUAMOSA promoter-binding-like transcription factors and microRNA156 in rice. *Plant Physiol* 142:280–293
- Xie J (2013) Study on the florescence biology of *Carya illinoensis*. Dissertation, Nanjing Forestry University
- Xing S, Salinas M, Höhmann S, Berndtgen R, Huijser P (2010) Mir156-targeted and nontargeted sbp-box transcription factors act in concert to secure male fertility in arabidopsis. *Plant Cell* 22:3935–3950
- Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T, Yabuki T, Aoki M, Seki E, Matsuda T, Nunokawa E, Ishizuka Y, Terada T, Shirouzu M, Osanai T, Tanaka A, Seki M, Shinzoka K, Yokoyama S (2004) A novel zinc-binding motif revealed by solution structures of DNA-binding domains of arabidopsis SBP-family transcription factors. *J Mol Biol* 337:49–63
- Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591
- Yang Z, Wong WS, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 22:1107–1118
- Zhang B, Xia GH, Wang ZJ, Huang JQ (2008) Studies on flowering biological characteristics and receptive period of female flowers of *Carya cathayensis*. *J Southwest For Coll* 28:1–4
- Zhang LS, Wu B, Zhao DG, Li CL, Shao FJ, Lu SF (2014) Genome-wide analysis and molecular dissection of the SPL gene family in *salvia miltiorrhiza*. *J Integr Plant Biol* 56:38–50

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.