INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY
ISSN Print: 1560–8530; ISSN Online: 1814–9596
19–1357/2020/23–5–951–962
DOI: 10.17957/IJAB/15.1374
http://www.fspublishers.org

Full Length Article

Cloning and Expression Features of Flowering Integrator Genes *ShSOC1-1* and *ShSOC1-2* from Chia (*Salvia hispanica*)

Bao-Jun Chen[†], Yu-Fei Xue[†], Xue-Li Liao, Cheng-Long Yuan, Lin Zhang, Xia Yang, Man-Lin Jiang, Yan-Xi Qu and You-Rong Chai[*]

Chongqing Key Laboratory of Crop Quality Improvement; Chongqing Rapeseed Engineering Research Center; Engineering Research Center of South Upland Agriculture of Ministry of Education; Academy of Agricultural Sciences, Southwest University; College of Agronomy and Biotechnology, Southwest University, Tiansheng Road 2#, Beibei, Chongqing, 400715, China

[*]For correspondence: chaiyourong@163.com; 2285105844@qq.com; 1056479422@qq.com; chaiyourong@163.com

[†]Contributed equally to this work and are co-first authors

Received 02 September 2019; Accepted 23 December 2019; Published 03 March 2020

Abstract

Chia (*Salvia hispanica*) is an ancient Mexican Indian crop with outstanding health-promoting values, and is reviving in recent years, but the short-day habit is a crucial limitation for its worldwide spreading. The MADS-box protein SOC1 can integrate pathways signals to promote flowering, hence this study isolated the chia *SOC1* subfamily, and characterized its molecular, evolutionary and expression features. *SOC1* was originated along the origin of seed plants, and *SOC1* gene number was amplified independently in gymnosperms and monocots at order/family scales, while all eudicots have 3 *SOC1* genes. The full-length cDNAs of *ShSOC1-1* and *ShSOC1-2* are 1103–1299 and 1003 bp, orthologous to *Arabidopsis thaliana ALG20*/*SOC1* and *AGL42/71/72*, respectively. *ShSOC1-1* expression is highest in stems and young buds, followed by mature leaves, semi-mature buds, mature buds, flowers and young leaves. *ShSOC1-2* expression is highest in young buds, but much lower in other organs. They both show lowest expression in seeds. *ShSOC1-1* is abundant at sunny night and is suppressed by rainy weather, whereas *ShSOC1-2* is suppressed in hot summer and is abundant in rainy afternoon and autumnal equinox midnight. In short-days, *ShSOC1* is down-regulated by brassinolide (BR), kinetin (KT) and indole-3-acetic acid (IAA) but upregulated by gibberellin (GA_3), *ShSOC1-2* being more sensitive. In long-days, *ShSOC1* is downregulated by BR but upregulated by KT, IAA and GA_3, *ShSOC1-1* being more sensitive. Cold stress upregulated *ShSOC1-1* but suppressed *ShSOC1-2*. Under heat, NaCl and abscisic acid (ABA) treatments, they both show fluctuant downregulation. They both might encode positive regulators of flowering, *ShSOC1-1* being basic, while *ShSOC1-2* being a threshold regulator. Our results shed light on photoperiodic and hormonal regulation of SOC1-associated flowering of chia and other short-day plants. © 2020 Friends Science Publishers

Keywords: Abiotic stresses; Chia; Evolution; Expression; Photoperiod; Phytohormones; *SOC1*

Introduction

Alpha-linolenic acid (ALA, n:3/omega-3) and linoleic acid (LA, n:6/omega-6) serve as precursors for the synthesis of long chain fatty acids such as eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) (Tang *et al.* 2018). EPA and DHA can not only promote brain development, improve visual acuity, prolong life expectancy, but also help in eradicate heart disease, hypertension, diabetes and cancer (Bai *et al.* 2016). Since the rediscovery of ancient and sacred oil crop chia, it has become more and more attractive due to its high unsaturated fatty acid content such as ALA and other important nutraceuticals such as proteins, edible fiber, seed coat gel, vitamin E, and flavonoid antioxidants (Sreedhar *et al.* 2015). Among known crops, Chia as an oil plant contains the highest ALA level of the total fatty acids. It is grown in deserts below an altitude of 4,000 feet in Mexico and Southwest America, and has developed into one of the important staple crops by ancient Astek and Mayas (Ayerza and Coates 2005).

Plant flowering time is mainly regulated by five pathways: photoperiod pathway, vernalization pathway, gibberellin pathway, autonomous pathway and aging pathway. The MADS-box transcription factor *SOC1* (AGL20) is an integrator that integrates flowering signals from multiple flowering regulatory pathways to promote flowering (Richter *et al.* 2013). Functional loss in *Arabidopsis thaliana SOC1* mutants or transgenic silencing of *SOC1* led to late flowering. When *A. thaliana SOC1*-like genes were transformed into *Citrus reticulate*, flowering time was significantly accelerated (Tan and Swain 2007). Transforming petunia *SOC1*-like/*FBP21* into tobacco resulted in early flowering, without influence on flower number and seed quality (Ma *et al.* 2011).

To cite this paper: Chen BJ, YF Xue, XL Liao, CL Yuan, L Zhang, X Yang, ML Jiang, YX Qu, YR Chai (2020). Cloning and expression features of flowering integrator genes *ShSOC1-1* and *ShSOC1-2* from chia (*Salvia hispanica*). *Intl J Agric Biol* 23:951–962

Surprisingly, overexpression of *Gossypium hirsutum GhSOC1*-like that is orthologous to *AtAGL42/71/72* in *Gerbera jamesonii* did not advance the flowering time, but interfered with floral organ development (Ruokolainen *et al.* 2011). Rice *FDRMADS8*, which is an orthologue of *AtAGL14*, has certain influence on floral development (Jia *et al.* 2000) and *AtAGL14* also promotes root development in *Arabidopsis* (Garay-Arroyo *et al.* 2013). *AtAGL19* promoted early flowering in *hda9* mutants, and was inhibited in *HDA9* or in LDs (Kim *et al.* 2013). The photoperiod pathway enhances *SOC1* and promotes flowering by promoting the expression of *FT* and *CO* (Yoo *et al.* 2005). *FT* and *SOC1* are affected by the change of photoperiod from the leaf to the apical meristem, promoting the plant from vegetative to reproductive growth (Immink *et al.* 2012). Autonomous and vernalization pathways inhibit the expression of flowering repressor *FLC* and then promote flowering. *FLC* delays flowering by delaying *SOC1* expression in meristems by inhibiting *FT* (Li *et al.* 2016a; Richter *et al.* 2019). The gibberellin pathway becomes the main flowering pathway that affects *SOC1*, rather than *FLC* and *FT* in SDs (Moon *et al.* 2003). In *Arabidopsis*, SOC1 and SOC1-like (AGL42, AGL71, AGL72) together constitute a subgroup of the MIKCC-type MADS-box transcription factors, and its main function is to participate in the control of the plants flowering time. Involved in the regulation of floral organ development, *SOC1* directly controls *SOC1*-like to balance these *SOC1*-like expression levels (Dorca-Fornell *et al.* 2011), and thus precisely controls flowering. The *SOC1* subgroup have been cloned in various plants, such as barley (Papaefthimiou *et al.* 2012), soybean (Zhong *et al.* 2012), mango (Wei *et al.* 2016), *Brassica juncea* (Li *et al.* 2019) *etc*. There is no report on cloning and research of *SOC1* subgroup genes in the Lamiales order.

According to the latest Chile climate simulation research, the suitable cultivation area of traditional short-day chia genotypes in China is restricted to mountainous regions of southern Yunnan and Taiwan (Cortés *et al.* 2017). This research group took the lead in carrying out the chia research in China. Since 2016, it was found that sowing in spring and summer in Beibei and Hechuan (30° N), Chongqing, China, chia bloomed in October, and a small amount of mature seeds could be harvested in late autumn. Though a small proportion of the seeds on the ear could mature, most seeds could not reach full maturity, indicating that the southern part of the Sichuan Basin, which is warm and frost-free in the winter, is a planting northern margin of traditional chia. This means that if flowering time is promoted a little, chia can be grown not only in China's low-latitude climate-friendly regions but also in the entire Sichuan Basin and many other parts of China's mid-latitudes. Even for the whole world, short-day photoperiod habit and late flowering of chia is the crucial limiting factor for its crop spreading, thus dissecting chia flowering mechanism has become a necessary basic work. In this study, two *SOC1* genes were cloned from chia, the basic characteristics of genes and proteins were analyzed, the plant *SOC1* evolutionary characteristics were revealed, and their transcriptional organ-specificity and responsiveness to multiple hormones under long/short photoperiods, circadian rhythms, seasonal transitions and abiotic stresses were also investigated. It will promote the understanding of the chia flowering mechanism, and enrich the understanding of *SOC1*.

Materials and Methods

Plant materials, treatment, and nucleic acid extraction

Chia was grown at Hechuan Farm, Southwest University, sown on May 24, 2016. On August 21–22, September 5–6, September 20–21 and October 5–6, mature leaves were collected at 2:58, 5:58, 9:28, 12:58, 16:28, 19:58, 23:28 of the day. They were used for gene cloning and to detect diurnal styles of gene expression. Root (Ro), stem (St), young leaves (YL), mature leaves (ML), young buds (YB, about 5 days old), semi-mature buds (SMB, about 10 days old), mature buds (MB, about 15 days old), flowers (Fl), early-stage seeds (ES, about 10 days old), middle-stage seeds (MS, about 20 days old) and late-stage seeds (LS, about 30 days old) were sampled for detecting organ-specificity of the cloned genes.

The methods used to cultivate the seedlings of chia in the artificial climate chamber referred by Xue *et al.* (2017). The six-leaf stage seedlings were moved to the plant growth chamber for treatment with 2 styles of photoperiods. The long-day treatment was 16 h-light and 8h-dark, and the short-day treatment was 12h-day and 12-night, with constant temperature of 30ºC and relative humidity of 56%. Each photoperiod treatment lasted for one week. Four hormone treatments were carried out, *i.e.* 80 μmol L^{-1} kinetin (KT), 2 μmol L^{-1} brassinolide (BR), 200 μmol L^{-1} gibberellin (GA_3) and 250 μmol L^{-1} indole acetic acid (IAA). Each hormone was treated for 0 d (control/CK, basal level), 1 d, 3 d and 9 d. Mature leaves were sampled at each time point for characterization of phytohormone responsiveness of *ShSOC1-1* and *ShSOC1-2*.

Chia seedlings were cultured in the artificial climate chamber and subjected to high temperature at 38ºC, low temperature at 4ºC, mechanical wounding, 100 μmol L^{-1} methyl jasmonate (MeJA), 100 μmol L^{-1} abscisic acid (ABA), 1 mmol L^{-1} salicylic acid (SA), 300 mmol L^{-1} sodium chloride (NaCl) and 10% polyethylene glycol 6000 (PEG6000). At 0 h, 0.5 h, 3 h, 9 h, 24 h, and 48 h time points after treatment, mature leaf samples were taken for characterization of stress responsiveness of cloned genes (Xue *et al.* 2017).

Each study had 3 biological replicates. Samples were kept in liquid nitrogen for transportation and stored at -80°C. Total cellular RNA was extracted using the Biospin Plant Total RNA Extraction Kit (BioFlux, China), and total gDNA was extracted from mature

leaves using general CTAB method. Electrophoresis and spectrophotometric detection were adopted to detect the quality and quantity of the nucleic acids.

Cloning of the conservative sequences of *SOC1* genes from chia

In order to clone the conserved region of chia *SOC1* genes, the *A. thaliana SOC1* mRNA was retrieved from NCBI GenBank (NM_130128.4), and used as an electron probe for the *in silico* cloning the orthologous sequence from chia-relative species *Sesamum indicum* (sesame), *Erythranthe guttatu*, *Salvia pomifera* and *S. miltiorrhiza*, since there was no chia sequence in the GenBank. All *SOC1* annotation mRNA, TSA, EST and gDNA tag sequences were downloaded and multiple sequence alignment was performed. According to the conservative sites of *SOC1* alignments, degenerate primer combination *FLSOC1C* + *RLSOC1C* was designed (Table 1). One μg of total RNA mixed from all organs was subjected to gDNA deletion and reverse-transcription using the PrimeScript Reagent Kit with gDNA Eraser (TaKaRa Dalian, China) to obtain the first-strand total cDNA. Then it was used as a template for amplification of the conservative regions of chia *SOC1* genes using conventional PCR (Annealed at 58ºC and extended for 2 min). Conventional electrophoresis, gel recovery, pMD19-T vector (TaKaRa Dalian, China) recombination and *Escherichia coli* DH5α transformation were performed. After PCR test for positive clones, batches of clones corresponding to insert length polymorphism were sent to Shanghai Lifei Information & Technology Company (China) for sequencing using M13F/M13R primers.

5-RACE and 3-RACE of chia *SOC1* genes

The results of sequencing showed that the conservative regions of 2 chia *SOC1* genes were obtained, which were named as *ShSOC1-1* and *ShSOC1-2* respectively. Then 5'-RACE and 3'-RACE primers of *ShSOC1-1* and *ShSOC1-2* were designed (Table 1), according to the conservative region sequences. One μg of total RNA of organ-mixture was used as start material for RACE handling using the SMARTer™ RACE Amplification Kit (Clontech, U.S.A.) to obtain the first-strand total cDNA template of the 5'-RACE and 3'-RACE, respectively. Primers FShSOC1-13-1/FShSOC1-23-1 and FShSOC1-13-2/FShSOC1-23-2 were used for pairing with the universal primer LUPM and NUP (Table 1) for 3'-RACE primary and nested amplifications of *ShSOC1-1/ShSOC1-2*, respectively. The PCR annealing temperature was 63°C, and the extension time was 1 min. Primers RShSOC1-15-1/RShSOC1-25-1 and RShSOC1-15-2/RShSOC1-25-2 were matched with the universal primers LUPM and NUP (Table 1) for primary and nested amplifications of 5'-RACE of *ShSOC1-1*/*hSOC1-2*, respectively. The PCR annealing temperature was 62°C and the extension time was 1 min. Electrophoresis, gel recovery, TA cloning and sequencing were performed.

Cloning of full-length sequences of chia *SOC1* subfamily genes

Based on the conservative regions and 5'-RACE and 3'-RACE results, we can obtain the full-length cDNAs of *ShSOC1-1* and *ShSOC1-2* using Vector NTI assemblage function. Based on this, we designed the primer combinations *FShSOC1-1* + *RShSOC1-1* and *FShSOC1-2* + *RShSOC1-2* (Table 1). The two full-length cDNAs were amplified by PCR using 3'-RACE template, annealed at 55°C and extended for 2 min. Electrophoresis, gel recovery, TA cloning and sequencing were performed.

Bioinformatics analysis

GenBank sequence search, BLAST, *in silico* cloning, and CDD detection were performed at NCBI (http://www.ncbi.nlm.nih.gov). Vector NTI Advance 11.5.1 and DNAStar version 7.1.0 software were used for sequence creation, analysis, annotation, translation, comparison, assembly and other analysis. Protein analysis were performed at Expasy (http://www.expasy.org), GSDS2.0 (http://gsds.cbi.pku.edu.cn/), CBS (http://www.cbs.dtu.dk/Services/). According to the NCBI BLASTp chia SOC1 subfamily results, completed genome sequencing and representative species in plant taxonomy were selected, and then all their SOC1 proteins sequences were electronically cloned, and then multiple comparisons were performed using ClustalX V2.0 to generate fst files. SeaView 4.0 uses the muscle pattern to perform multiple comparisons. Under Distance Method and BioNJ method, Distance=Poisson and Bootstrap=1000 are set to build the phylogenetic tree and display the tree in Squared format.

qRT-PCR detection of transcript expression of chia *SOC1* genes

The transcriptional expression of *ShSOC1-1* and *ShSOC1-2* was detected by using *FShSOC1-1RT* + *RShSOC1-1RT*, *FShSOC1-2RT* + *RShSOC1-2RT* primer pairs, respectively. The *25SrRNA* was detected by *F25SRT* + *R25SRT* as internal control (Table 1). qRT-PCR was performed on a CFX Connect™ Real-Time PCR Detection System (Bio-Rad, U.S.A.) with a program of 95°C for 10 min, and 45 cycles of amplification (95°C for 10 sec, 60°C for 20 sec, 72°C for 10 sec). When qRT-PCR was completed, the temperature was raised from 65°C to 95°C and the melting curve was detected to confirm the specificity of the amplification.

Results

Cloning of full-length cDNAs of *ShSOC1-1* and *ShSOC1-2* genes

Electrophoresis analysis of PCR product of amplification for the conservative sequences of the chia *SOC1* genes

Table 1: Primers used in cloning and qRT-PCR detection of *SOC1* genes from Chia

Primers name	Primers sequence (5’→3’)	Application
FLSOC1C	AGAAATGGGCTGYTGAAGAARGC	Forward primer for Chia *SOC1* conservative regions amplification
RLSOC1C	GGBRGNCCDATGAACAATTCNGTCDCNAC	Reverse primer for Chia *SOC1* conservative regions amplification
FShSOC1-13-1	TTGAGCGCAGTGTCACCACCATTCGT	GSP for *ShSOC1-1* 3'-RACE primary amplification
FShSOC1-13-2	TTGGACTTCAAACACAAGGTGGAGG	GSP for *ShSOC1-1* 3’-RACE nested amplification
FShSOC1-23-1	TCCAGCGAAGCCTACACAATGTCAGG	GSP for *ShSOC1-2* 3’-RACE primary amplification
FShSOC1-23-2	GTGAAGTTAGGGAAACAGAAAGAGAGAG	GSP for *ShSOC1-2* 3'-RACE nested amplification
RShSOC1-15-1	CTGCATATTATGCTCCGAAGGTGGATTG	GSP for *ShSOC1-1* 5'-RACE primary amplification
RShSOC1-15-2	TGAGCTTGCAAATTCATGGAGCTTGCC	GSP for *ShSOC1-1* 5’-RACE nested amplification
RShSOC1-25-1	CTTCATGCGTTGTTCGACTTCATTGCC	GSP for *ShSOC1-2* 5’-RACE primary amplification
RShSOC1-25-2	TTGGAGCTTGAGAACTCATAAAGTCTTCC	GSP for *ShSOC1-2* 5’-RACE nested amplification
LUPM	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT	Anchor primer for 5'-and 3'-RACE primary amplification
NUP	AAGCAGTGGTATCAACGCAGAGT	Anchor primer for 5'-and 3'-RACE nested amplification
FShSOC1-1	ACTGTGAATATTACTCCTAGTACTACTAC	*ShSOC1-1* full-length forward primer
RShSOC1-1	AATGCATAAAAAAGTTGTCATTAGTTAATAAATA	*ShSOC1-1* full length reverse primer
FShSOC1-2	CCTCCCTCTCTCTCTCTCTCTCTCACAT	*ShSOC1-2* full-length forward primer
RShSOC1-2	AGATTAAAGATGCATCCAAAAGAATTTCCAG	*ShSOC1-2* full length reverse primer
F25SRT	GATTTCTGCCCAGTGCTCTGAA	*25SrRNA* qRT-PCR forward primer
R25SRT	TCTGCCAAGCCCGTTCCCTT	*25SrRNA* qRT-PCR reverse primer
FShSOC1-1RT	GGCTACTTGGTGAAGGGTTAGG	*ShSOC1-1* qRT-PCR forward primer
RShSOC1-1RT	CTCTCCTCGTTCGATCCTCCT	*ShSOC1-1* qRT-PCR reverse primer
FShSOC1-2RT	GCAATGAAGTCGAACAACGCAT	*ShSOC1-2* qRT-PCR forward primer
RShSOC1-2RT	CATTGTGTAGGCTTCGCTGGA	*ShSOC1-2* qRT-PCR reverse primer

showed a 0.6-kb specific band. Sequencing of 20 positive clones produced 2 member genes and NCBI BLASTn showed orthologs to the *SOC1* (*AtAGL20*) and *SOC1*-like (*AtAGL42*) of plants, and they were named as *ShSOC1-1* and *ShSOC1-2*, respectively. No significant band was found in the primary amplifications of 5'-RACE and 3'-RACE of *ShSOC1-1* and *ShSOC1-2*, with only smear at the predicted size. The 5'-RACE nested PCRs of *ShSOC1-1* and *ShSOC1-2* each generated a band of about 350 bp. After TA cloning, 5'-RACE clones had insert length polymorphisms. The net length of *ShSOC1-1* clones after batch sequencing was 206, 231, 418, 377 bp, with intron retention in some clones. The net length of *ShSOC1-2* clones was 251, 286, 309, and 310 bp. The 3'-RACE of *ShSOC1-1* and *ShSOC1-2* nested PCR generated 2 bands of about 0.35 kb and 0.45 kb, respectively. All the 3'-RACE clones had polymorphic insert length after TA cloning. The net *ShSOC1-1* clones were 298, 344, 349, and 397 bp, with a net length of 369, 430, 435, 461, and 485 bp for *ShSOC1-2* (Poly A not included). Based on the RACE results, about 1.2 kb and 1 kb band identical to the expected size was obtained by amplifying the full-length cDNA of *ShSOC1-1* and *ShSOC1-2* using end-to-end PCR primer combinations. The sequences corresponded to the assembled ones. Sequence analysis revealed intron retention in the 5'-RACE of some of the mRNA molecules of *ShSOCl-1*, so *ShSOCl-1* has 3 versions of mRNA, whereas *ShSOCl-2* has only one version of mRNA. We chia total gDNA was used as template to amplify the full-length gDNAs of *ShSOC1-1* and *ShSOC1-2*, which was failed even we replaced reagents such as enzymes and optimized the amplification cycle parameters, indicating that they either have very long introns or have very complex structures.

Structure and features of *ShSOC1-1* and *ShSOC1-2* genes

Fig. 1 shows that *ShSOC1-1* has 3 versions of mRNA (GenBank Accession Numbers MF577048, MF577049 and MF577050). The longest standard mRNA is 1103 bp (Poly A not included), the longest mRNA of 5'-UTR intron-retention is 1299 bp, and the longest mRNA is 1112 bp which has a 9-bp alternative splicing at the right border of the second intron. The longest mRNA of *ShSOC1-2* is 1003 bp (Poly A not included, GenBank Accession Number MF577051). The normal 5′ -UTR, ORF and 3′ -UTR of *ShSOC1-1*/*ShSOC1-2* are 196/128 bp, 666/642 bp, and 241/233 bp, respectively. The G+C content of 5′ -UTR, ORF, and 3′ -UTR of *ShSOC1-1*/*ShSOC1-2* were 36.2/41.4%, 52.0/44.8% and 29.5/29.2%, respectively. There were 3 and 4 transcription initiation sites for *ShSOC1-1* and *ShSOC1-2*, with 4 and 6 poly A tail sites, respectively. The identity percentages between *ShSOC1-1* and *ShSOC1-2* on mRNA and ORF levels were 54.1 and 58.7%, respectively. BLASTn showed that *ShSOC1-1* has higher homology with sesame *SOC1*, followed by *E. guttata SOC1*; and *ShSOC1-2* has higher homology with sesame *AGL42*, followed by *E. guttata SOC1*-like. The phylogenetic tree of coding regions indicates that *ShSOC1-1* and *ShSOC1-2* are orthologous to the *SOC1* and *SOC1*-like in the plant kingdom, respectively.

Characterization of deduced *ShSOC1-1* and *ShSOC1-2* proteins

The *ShSOC1-1* and *ShSOC1-2* proteins are 221 aa and 213 aa in length, with theoretical MWs of 25.04 and 24.97 kD,

```
   1 ACTGTGAATATTACTCCTAGTACTACTACTACTATCCTCTCTGTAAAACCATTATTATTATTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC
 101 TCTCTCATAAACATAAACACACCATATTTGTTTCTTCCCTTTTTTAGAACCTCAAACCAAAAGTCTCTCTGGGTGTTTGCCATATTTGTAAGCATCCCTT
 201 CTTGTTTCCTCAAAAAACCCACTTATTTTCGGAAATTCTTTGCAAGACCTTTCTTTTTTGCAGAAGAGAGTAATCCAAATACTTTCTACATATTCACTCT
 301 CACACAGCAACATTCCTCTCAAAGATTTGATTTTTTTCCACTTGCTCTCCATAAAAATTTATTTCTTTGACAAGGGCTTGAAGATGGTGAGAGGGAAGAC
                                                                                      M  V  R  G  K  T
 401 TCAGATGCGGCGTATTGAGAACGCCACAAGCAGGCAAGTGACCTTCTCCAAGAGGAGAAATGGGCTGCTCAAGAAGGCCTTTGAGCTCTCAGTTCTCTGC
      Q  M  R  R  I  E  N  A  T  S  R  Q  V  T  F  S  K  R  R  N  G  L  L  K  K  A  F  E  L  S  V  L  C
 501 GATGCTGAGGTGGCCCTCATCATCTTCTCCCCTAGGGGCAAGCTCCATGAATTTGCAAGCTCAAGCATGCACGAGACGATCGAGCGTTACCAGAAGTATA
     D  A  E  V  A  L  I  I  F  S  P  R  G  K  L  H  E  F  A  S  S  S  M  H  E  T  I  E  R  Y  Q  K  Y  T
 601 CTAAAGATAGCCAAGCCAACAATCCACCTTCGGAGCATAATATGCAGCAATTGAAGCACGAAGCTGCAAGCATGATCAAAAAAATCGAGCAGCTGGAATC
       K  D  S  Q  A  N  N  P  P  S  E  H  N  M  Q  Q  L  K  H  E  A  A  S  M  I  K  K  I  E  Q  L  E  S
 701 ATCAAAGCGGAGGCTACTTGGTGAAGGGTTAGGGCCGTGCACCATCGAGGAGCTGCAGCAGCTCGAGCAGCAGCTTGAGCGCAGTGTCACCACCATTCGT
      S  K  R  R  L  L  G  E  G  L  G  P  C  T  I  E  E  L  Q  Q  L  E  Q  Q  L  E  R  S  V  T  T  I  R
 801 GCACGAAAGATGCAAGTGTACAAGCAGCAGATCGAGCATTTAAAGGAAAAGGAGAAAGCCCTAGCTGCTGAAAATGCAATGCTGTGTGAGAAGGGGTTGC
     A  R  K  M  Q  V  Y  K  Q  Q  I  E  H  L  K  E  K  E  K  A  L  A  A  E  N  A  M  L  C  E  K  G  L  Q
 901 AGTTTGGACTTCAAACACAAGGTGGAGGGGGAGGAGGAGGATCGAACGAGGAGAGAGCAGTGGTGCCATCTGCGGAGATCAGTGAGGTGTCGGATGTGGA
       F  G  L  Q  T  Q  G  G  G  G  G  G  G  S  N  E  E  R  A  V  V  P  S  A  E  I  S  E  V  S  D  V  E
1001 GACGGAGTTGTTCATAGGGCTGCCTGAGACGAGGCATAAGCGCCCCTACCAGAAATAGTATGTGAATAAAGTTAATTAAAGCTAGCCAACTTCCTAGCTT
      T  E  L  F  I  G  L  P  E  T  R  H  K  R  P  Y  Q  K  *
1101 TAGTAACTTATGTCTGTATATATACATATTATTGCTTATTTTATTTTTTATGTTTGCACCTAGAAGCTGCAAATAATTCGTGCCTACTGCATCTGTTAAA
1201 CCACATTTGTTTCGTTTTGGCTGTTGATGGAATGTGGTATTTTTCAGATCATGGTTCTGATGAATTATTTATTAACTAATGACAACTTTTTTATGCATTa
1301 aaaaaaaaaaaaaaaaaaaaaaaa
```

ShSOC1-1

```
   1 CCTCCCTCTCTCTCTCTCTCTCTCACATCACATGTTTCTCTATCTGGGATTCATGGGAGTTCATCCTTTTTCCTTTTTGTAAAACTTTCCCACAATCAAA
 101 TTGATTGATAGCCAGAAAGAGGGGGAAAATGGTGAGAGGGAAGGTGGAAATGAAGCGCATTGAAAATCCAACAAGCAGGCAGGTGACCTTTTCAAAGCGC
                                 M  V  R  G  K  V  E  M  K  R  I  E  N  P  T  S  R  Q  V  T  F  S  K  R
 201 AGAAACGGGCTATTGAAGAAGGCATATGAGCTCTCAGTGCTCTGCGATGCAGAAGTTGCTTTGATCATATTCTCTCAAAAAGGAAGACTTTATGAGTTCT
     R  N  G  L  L  K  K  A  Y  E  L  S  V  L  C  D  A  E  V  A  L  I  I  F  S  Q  K  G  R  L  Y  E  F  S
 301 CAAGCTCCAACATGCAGAAGTCCATAGAGAAATATCTTGAACGTGCAAGAGAAGAGCGGGGCAATGAAGTCGAACAACGCATGAAGTTGCAGTATTTAAA
       S  S  N  M  Q  K  S  I  E  K  Y  L  E  R  A  R  E  E  R  G  N  E  V  E  Q  R  M  K  L  Q  Y  L  K
 401 GCATGAAGCAGCGTTGATGATGAAGAAGATAGAGCTACTAGAGAATTCCCAAAGGAAGCTTTTAGGGCAAAACCTTGGAGCTTATTCATTGCAAGAGCTC
      H  E  A  A  L  M  M  K  K  I  E  L  L  E  N  S  Q  R  K  L  L  G  Q  N  L  G  A  Y  S  L  Q  E  L
 501 CACGAAACCGAAAACCAGCTCCAGCGAAGCCTACACAATGTCAGGGCCAGAAAGGCTCAACTATACAAGGAGGAAACAGAAAAACTACTTGCAAAGGAGA
     H  E  T  E  N  Q  L  Q  R  S  L  H  N  V  R  A  R  K  A  Q  L  Y  K  E  E  T  E  K  L  L  A  K  E  K
 601 AGTTTTTGTTGGAAGAAAATTCCAAGTTGCGCCAAAAGTGTGAAGTTAGGGAAGCAGAAAGAGAGAGTCAGAGCCAAACCACACTGAGCTCGGAGGTTGT
       F  L  L  E  E  N  S  K  L  R  Q  K  C  E  V  R  E  A  E  R  E  S  Q  S  Q  T  T  L  S  S  E  V  V
 701 GACCGAGTTGTTCATCGGCCCACCAACTACAAGGAAAGCCCGTGATTTCTCAACTTTTACTAGCAAATAAAAATGCTTTTTTTTTCTTATGGCATACATA
      T  E  L  F  I  G  P  P  T  T  R  K  A  R  D  F  S  T  F  T  S  K  *
 801 TAGACAAGTATATTTATCGTAAATATATACGTTAAACAAATAAGTGTATTTTGTTGTTCAGTATATATAGTGGGCCATTAGAACTAGCAAAATAAAAATG
 901 TAATGGGATATTTCATTGTTCAATACTAGTATAAGGGATTGGACATCCTAAGATAAATTGGGCCATTGTGGGCTGGAAATTCTTTTGGATGCATCTTTAA
1001 TCTaaaaaaaaaaaaaaaaaaaaaaaaaa
```

ShSOC1-2

Fig. 1: mRNA and encoded amino acid of *ShSOC1-1* and *ShSOC1-2* genes

The start codon ATG and the stop codon TGA/TAG are in underlined bold face. Transcription start sites and polyadenylation sites are in underlined and italic bold face. The possible polyadenylation signals are in italic bold face. 5'UTR-intron in *ShSOC1-1* is labeled with grey background. Poly A tail is in lower case

pIs of 9.06 and 9.55, respectively, and which are alkaline. The identity percentage between *ShSOC1-1* and *ShSOC1-2* is 50.2% and the positives percentage is 61.8%. BLASTp showed that *ShSOC1-1* has higher homology with *E. guttata* and sesame *SOC1*, followed by *Plantago major SOC1*; *ShSOC1-2* has higher homology with the sesame *SOC1* and *AGL42*, followed by the *E. guttata SOC1*.

NCBI BLASTp CDD analysis showed that ShSOC1-1 and ShSOC1-2 have MADS and Coiled coil domains. SignalP 4.1 predicted that ShSOC1-1 and ShSOC1-2 do not contain a signal peptide. BaCelLo (http://gpcr.biocomp.unibo.it/bacello/pred.htm), EpiLoc (http://epiloc.cs.queensu.ca/), Plant-mPLoc (www.csbio.sjtu.edu.cn/bioinf/plant-multi/), and YLoc (www.multiloc.org/YLoc) predicted *ShSOC1-1* and *ShSOC1-2* are located in the nucleus. SLP-Local predicts that they are located to the mitochondria or nucleus. Taken together, they are most probably located to the nucleus, consistent with the identity of the MADS-box transcription factors. TMHMM2.0 (www.cbs.dtu.dk/services/TMHMM/) and TOPCONS (http://topcons.net/) predicted that *ShSOC1-1* and *ShSOC1-2* have no transmembrane domain. NetPhos3.1 (www.cbs.dtu.dk/services/NetPhos/) predicted that *ShSOC1-1* has 23 potential phosphorylation sites, including 15 S (serine), 7 T (threonine), and 1 Y (tyrosine), and *ShSOC1-2* have 21 potential phosphorylation sites, with 11 S, 9 T, and 1 Y.

In the secondary structures predicted by SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/secpred_sopma.pl), α-helix of *ShSOC1-1/ShSOC1-2* as high as accounted for 59.73/68.08%, β-sheet (extended strand) accounted for 7.69/14.08%, β-turn accounted for 7.24/5.16%, and random coil accounted for 25.34/12.68% (Fig. 2). *ShSOC1-1* has several large α-helix in middle sites, while *ShSOC1-2* has a very large α-helix throughout the middle sites and the α-helix occupies a large proportion in these two genes. However, the N-terminal α-helix of *ShSOC1-1* is apparent, whereas the N-terminus of *ShSOC1-2* has no α-helix.

The tertiary structures of *ShSOC1-1* and *ShSOC1-2* were predicted by SWISS-MODEL (https://swissmodel.expasy.org), which are similar to each other. In many flowering plants, the *SOC1* genes are important flowering regulatory genes, known as the signal integrator of flowering pathway (Chen *et al.* 2017). In combination with the protein structures, key sites in the conserved region, and physicochemical properties of *ShSOC1-1* and *ShSOC1-2*, it was concluded that *ShSOC1-1* and *ShSOC1-2* genes may be involved in the regulation of flowering traits in Chia.

The phylogenetic relationship of *ShSOC1* proteins and *SOC1* evolution in plant kingdom

In order to systematically explore the phylogenetic relationships of the *ShSOC1* genes and reveal some new features of the evolution of the plants *SOC1* subgroup in the plant kingdom, we selected different taxa from the plant kingdom (chlorophytas, ferns, gymnosperms, monocots, and dicots) with whole genomes being sequenced to carry out phylogenetic analysis. Their *SOC1* proteins were electronically cloned and then *ShSOC1-1* and *ShSOC1-2* were used together to construct a phylogenetic tree of *SOC1* proteins (Fig. 3). The phylogenetic relationships among the species showed by this tree conform to taxonomy positions established by traditional community. Some new

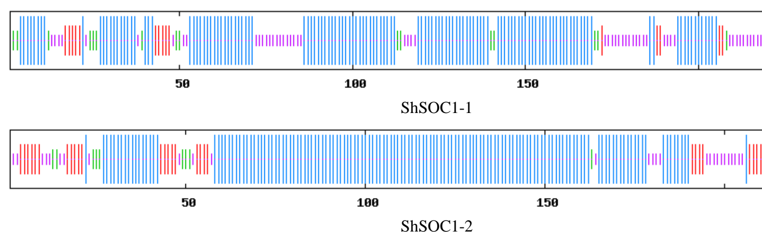

Fig. 2: Predicted secondary structures of *ShSOC1-1* and *ShSOC1-2*

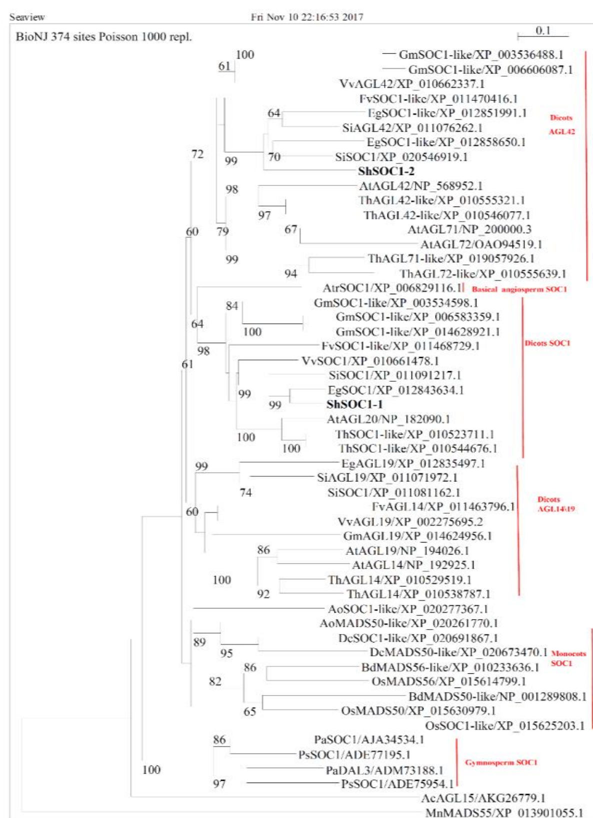

Fig. 3: Phylogenetic relationship of *SOC1* subgroup proteins from plant kingdom

Ac, *Adiantum capillus-veneris*; Ao, *Asparagus officinalis*; At, *Arabidopsis thaliana*; Atr, *Amborella trichopoda*; Bd, *Brachypodium distachyon*; Dc, *Dendrobium catenatum*; Eg, *Erythranthe guttatus*; Fv, *Fragaria vesca*; Th, *Tarenaya hassleriana*; Gm, *Glycine max*; Mn: *Monoraphidium neglectum*; Os: *Oryza sativa*; Pa: *Picea abies*; Ps: *Picea sitchensis*; Sh, *Salvia hispanica*; Si, *Sesamum indicum*; Th: *Tarenaya hassleriana*; Vv, *Vitis vinifera*

phenomena can be observed. There is no *SOC1* in non-seed plants (green aglae, ferns, *etc*.). Although there are MADS-box transcription factors in *Monoraphidium neglectum* and *Adiantum capillus-veneris*, there is a non-orthologous relationship with *SOC1*. The origins of *SOC1* is accompanied by the birth of sexual reproduction through “flowering” and seed setting.

The plant *SOC1* subfamily phylogenetic tree has two clusters: gymnosperms (*Picea abies*, *Picea sitchensis*) and angiosperms. The gymnosperm *SOC1* cluster occupies the basal sites, and the angiosperm *SOC1* cluster is divided into 4 branches. The monocots plant has 1 branch, and the dicots plant have 3 branches corresponding to *Arabidopsis SOC1/AGL20, AGL42/71/72/SOC1*-like, and *AGL14/19/SOC1*-like respectively. Within gymnosperms and monocots, it seems that *SOC1* duplication events widely occurred at order or family levels, resulting in 2 or more *SOC1* genes in most analyzed species. The ancestor of dicots has experienced *SOC1* twice duplication events, making the 3-*SOC1*-gene basic status of all eudicots. In Brassicales, more than one families have experienced their own *AGL14* and *AGL42* duplications, causing the *AGL14/19* double gene and *AGL42/71/72* triple gene status. At protein sequence level, *AGL42* is conserved, while *AGL71/72* diverged rapidly. In non-Brassicales eudicots, there is only one gene corresponding to Brassicales *AGL14/19* and *AGL42/71/72* respectively. Besides Brassicales *AGL71/72*, soybean *GmSOC1*-like/XP_003536488.1 and *GmSOC1*-like/XP_006606087.1, strawberry FvAGL14/XP_011463796.1, *Brachypodium distachyon* BdMADS50-like/NP_001289808.1, rice *OsSOC1*-like/XP_015625203.1 etc. also evolved rapidly at protein sequence level. It is noteworthy that the basal angiosperm *Amborella trichopoda* (common ancestor of monocots and dicots) has only one *SOC1* gene, indicating that angiosperms *SOC1* duplication events occurred after separation from basal angiosperms, and the non-duplicated and non-diversified status of important floral genes such as *SOC1* might be associated with the primitive and simple floral traits of basal angiosperms.

The phylogenetic tree also showed that *ShSOC1-1* and *ShSOC1-2* were orthologous to *A. thaliana* SOC1/AGL20 and *AGL42/71/72*, respectively. They are highly homologous to orthologous genes from *Salvia*, *S. indicum* and *E. guttata*, followed by other dicots plants. Since the orthologous gene corresponding to *AtAGL14/19* is common in dicots, it is speculated that this orthologous gene also exists in chia and needs to be cloned.

Organ specificity of *ShSOC1-1* and *ShSOC1-2* expression

The results of qRT-PCR showed (Fig. 4) that the expression of *ShSOC1-1* and *ShSOC1-2* in all organs was significantly different from each other. *ShSOC1-1* was highest in stems and buds, and high or significant in mature leaves, semi-mature buds, mature buds and flowers, but lowest in seeds. *ShSOC1-2* was highest in young buds and significantly lower in other organs, especially in seeds.

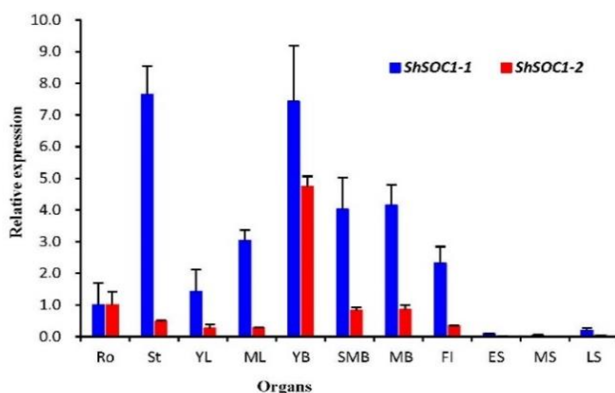

Fig. 4: Relative expression of *ShSOC1-1* and *ShSOC1-2* genes in different chia organs

Ro: root; St: stem; YL: young leaf; ML: mature leaf; YB: young bud; SMB: semi-mature bud; MB: mature bud; Fl: flower; ES: early-stage seed; MS: middle-stage seed; LS: late-stage seed

According to the relative quantitative preliminary judgments, the expression of *ShSOC1-1* in various organs is generally higher than that of *ShSOC1-2*, especially in stems, leaves, buds and flowers, implying that *ShSOC1-1* may be more effective than *ShSOC1-2* in regulating whole plant overall functions as well as the basal role during the flowering process. *ShSOC1-2* is mainly synergistically up-regulated at the key time-point for flowering determination.

Circadian rhythm of *ShSOC1-1* and *ShSOC1-2* expression and its response to long-short photoperiod seasonal shift

The qRT-PCR was used to examine the circadian rhythms of *ShSOC1-1* and *ShSOC1-2* in mature leaves and the response to the seasonal change from long to short photoperiod. The results showed that the expression of *ShSOC1-1* and *ShSOC1-2* was significantly different (Fig. 5). On August 21–22 (LD, sunny, 28–38 °C), *ShSOC1-1* was lower in the morning and did not change very much, upregulated from the afternoon and increased significantly at night, but fluctuated slightly at midnight; *ShSOC1-2* was lower throughout the day, and there was almost no significant fluctuation. On September 5–6 (LD, rainy, 20–24°C), affected by the rainy weather, *ShSOC1-1* was very low with less fluctuation within a whole day, while *ShSOC1-2* was low in morning, with a peak in afternoon, and then gradually down-regulated until maintaining a relatively stable low level. On September 20–21 (Autumnal equinox, sunny, 20–28°C), the two genes were very similar, with very weak and small changes during the daytime, but increased at night, peaked at midnight, and then decreased. On October 5–6 (SD, cloudy to overcast, 20–29°C), the two genes within a whole day were significantly similar to those of August 21–22 and September 20–21; *ShSOC1-1* was low during the daytime and high at night, peaked at midnight, and then gradually decreased; *ShSOC1-2* was relatively lower with little change within a whole day. Taken together, *ShSOC1-1* is characterized by predominant expression at nights on sunny days but suppressed by rainy weather, while *ShSOC1-2* is inhibited by high temperatures sunny days in summer, and dominantly expressed in rainy afternoon and in midnight of autumnal equinox (critical time for determining flowering and early floral differentiation).

Effect of phytohormones on the expression of *ShSOC1-1* and *ShSOC1-2* under long- and short-photoperiods

In the present study, KT, BR, GA_3, and IAA treatments were performed on 6-leaf chia seedlings under long/short-day conditions (LD/SD) respectively, and the expression of *ShSOC1-1* and *ShSOC1-2* was detected by qRT-PCR (Fig. 6). There are similarities and differences among hormones as well as between genes and photoperiods. After BR treatment, *ShSOC1-2* and *ShSOC1-2* were slightly inhibited under LDs or SDs, with *ShSOC1-2* being more sensitive. After KT treatment, *ShSOC1-1* and *ShSOC1-2* were significantly inhibited under SDs, with *ShSOC1-2* being more sensitive, and *ShSOC1-1* restored earlier. However, the expression trends of *ShSOC1-1* and *ShSOC1-2* were consistent under LDs, firstly upregulated and then fluctuated, but *ShSOC1-1* was more sensitive. After IAA treatment, *ShSOC1-1* and *ShSOC1-2* showed a slowly inhibition under SDs, but the two genes had the same tendency under LDs, firstly up-regulated and then fluctuated, with *ShSOC1-1* being more sensitive. After GA_3 treatment, *ShSOC1-1* had little effect on under SDs but *ShSOC1-2* was significantly upregulated; however, *ShSOC1-2* had little effect but *ShSOC1-1* was upregulated under LDs. Taken together, the expression of *ShSOC1* was down-regulated by BR, KT, and IAA under SDs, but upregulated by GA_3, and *ShSOC1-2* was more sensitive than *ShSOC1-1*. Under LDs, the expression of *ShSOC1* genes were down-regulated by BR and upregulated by KT, IAA and GA_3, and *ShSOC1-1* was more sensitive than *ShSOC1-2*.

Effect of abiotic stresses on the expression of *ShSOC1-1* and *ShSOC1-2*

The MADS-box transcription factor *SOC1* subfamily genes regulate plant growth as well as control flowering time and have important basic functions, but reports on abiotic stresses influence on their expression are limited. We used 5-week old chia seedlings for treatments with a variety of abiotic stresses and examined changes in the expression of *ShSOC1-1* and *ShSOC1-2* based on qRT-PCR (Fig. 7). After cold treatment at 4°C, *ShSOC1-1* was continuously upregulated within 48 h, while *ShSOC1-2* was quickly and significantly inhibited, then slowly restored to basal level at 48 h. After treatments with heat stress at 38°C and NaCl, *ShSOC1-1* and *ShSOC1-2* showed fluctuation and a rough down-regulation, though *ShSOC1-1* was transiently upregulated at 0.5 h under high temperature and at 9 h

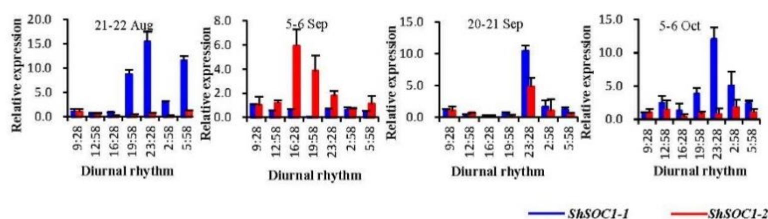

Fig. 5: Circadian rhythm of *ShSOC1-1* and *ShSOC1-2* expression, and response to seasonal change from long to short photoperiod

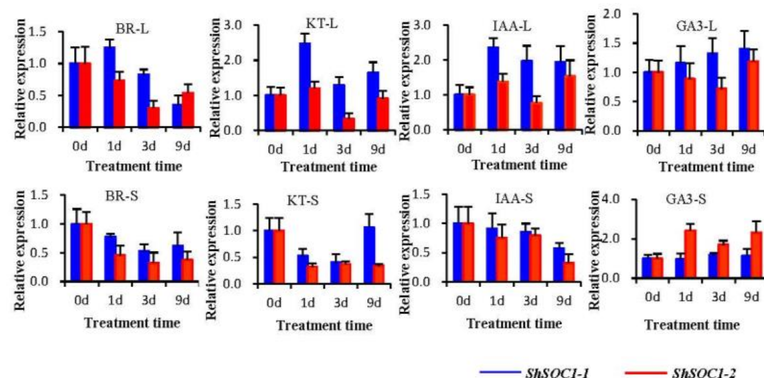

Fig. 6: Influence of phytohormones on the expression of *ShSOC1-1* and *ShSOC1-2* under long-photoperiod (-L) and short-photoperiod (-S) respectively

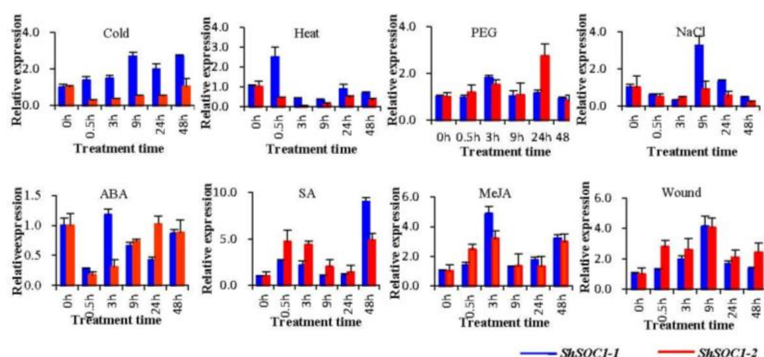

Fig. 7: Influence of abiotic stresses on the expression of *ShSOC1-1* and *ShSOC1-2*

under NaCl. ABA quickly inhibited their expression, but the recovery was also quick. Although there were small fluctuations in the expression of *ShSOC1-1* and *ShSOC1-2* after PEG treatment, the overall change was not clear. After MeJA, SA, and mechanical injury treatments, *ShSOC1-1* and *ShSOC1-2* showed a wavy up-regulation, especially when SA treatment was at 48 h, *ShSOC1-1* and *ShSOC1-2* were upregulated by 5–10 folds.

Discussion

In *A. thaliana*, all *SOC1* subfamily genes *AGL20/SOC1*, *AGL42*, *AGL71*, *AGL72*, *AGL14* and *AGL19* participate in the regulation of flowering, and are integrators of photoperiod, vernalization, gibberellin, autonomy, aging pathway and other flowering pathway signals, which occupy a central position in regulating flowering. *SOC1* is dominant within the subfamily, which regulates flowering time by regulating the same family members *AGL42*, *AGL71*, *AGL72*, *etc*. (Lee *et al.* 2000; Moon *et al.* 2003; Dorca-Fornell *et al.* 2011; Immink *et al.* 2012). The *Arabidopsis* flowering suppressor gene *FLC* normally inhibits expression of *SOC1* and rapidly upregulates *SOC1* in short-day *flc* mutants (Hepworth *et al.* 2002; Moon *et al.* 2003). Functional studies have shown that the function of the *SOC1* subfamily to promote flowering is at least conserved in angiosperms. In dicots, deletion of *Arabidopsis SOC1* and *SOC1*-like leads to late flowering (Hepworth *et al.* 2002). The overexpression of cabbage *BrAGL20* into *Brassica napus* resulted in early floral phenotype (Hong *et al.* 2013). *Mangifera indica MiSOC1* was transformed into *Arabidopsis*, which resulted in an early flowering stage (Wei

et al. 2016). *G. hirsutum GhSOC1* also promotes flowering after *A. thaliana* transformation, and it can regulate *APETALA1/FRUITFULL*-like gene *GhMADS42* to regulate flower organ morphology (Zhang *et al.* 2016). In monocot plants, overexpression of rice *OsSOC1* greatly advanced flowering time, *OsSOC1* was transformed to *Arabidopsis soc1* mutants to normalize flowering time (Andersen *et al.* 2004), and RNA interference of maize *ZmSOC1* gene postponed flowering time, while *ZmSOC1* overexpression or heterologous expression in *Arabidopsis* promoted early flowering (Alter *et al.* 2016). *Phyllostachys praecox SOC1* orthologous gene *PvMADS56* was transformed to *Arabidopsis* and the flowering time was also accelerated (Liu *et al.* 2016b). *Dendrobium nobile DnAGL19* gene transformed into *A. thaliana* promoted flowering through the *HOS1-FT* pathway (Liu *et al.* 2016c).

With a few exceptions, *FvSOC1* gene is an inhibitor of the flowering time in the perennial short-day plant wild strawberry, which regulates vegetative growth and reproductive growth respectively, through independent pathways (Mouhu *et al.* 2013). Overexpression of *G. hirsutum GhSOC1* in *G. jamesonii* does not advance the flowering time, but results in the decline of floral organs (Ruokolainen *et al.* 2011). *Actinidia chinensis SOC1*-like subfamily may not be involved in controlling the flowering time, but may affect the length of dormancy (Voogd *et al.* 2015). The I domain and C domain are lost in the protein encoded by *Kalanchoe daigremontiana KdSOC1* gene, and its function is to play an important role in the vegetative propagation of adventitious buds through the auxin signaling pathway, and its overexpression affects plant morphology (Liu *et al.* 2016a).

Chia *ShSOC1-1* and *ShSOC1-2* encode proteins with typical *SOC1* full domain features, which have also typical conserved domains and conserved sites. Their transcripts are highest in the early stage of flower bud differentiation. Given that higher plants *SOC1* subfamily generally regulates flowering and floral organ differentiation, it is speculated that the *ShSOC1* subfamily may be a positive regulator of chia flowering and early floral organ differentiation.

A large number of literatures have shown that phytohormones are involved in regulating flowering time, especially the gibberellin signaling pathway is one of the five major pathways of flowering induction, and the hormone pathway is intertwined with the photoperiodic and vernalization reactions (Shi *et al.* 2019). Although *SOC1* is one of the most important integrators of the five major pathways of flowering induction, there are not many reports on its response to the phytohormones. In view of this, four phytohormones were sprayed on chia plants under LDs and SDs in the present study. The results showed that under SDs *ShSOC1* subfamily was down-regulated by BR, KT and IAA, but upregulated by GA_3. *ShSOC1-2* was more sensitive than *ShSOC1-1*. *ShSOC1* subfamily was down-regulated under LDs, while *ShSOC1-1* was up-regulated by KT, IAA, and GA_3. *ShSOC1-1* was more sensitive than *ShSOC1-2*. *Arabidopsis SOC1* was slowly up-regulated after GA_3 treatment under both SDs and LDs (Moon *et al.* 2003), which is consistent with the trend in chia of this study. However, wheat *SOC1* and *SOC1*-like are down-regulated when treated with GA_3 under SDs (Pearce *et al.* 2013), which is contrary to the trend in chia of this study. It seems that low-temperature long-day plants and low-temperature short-day plants remain the same, and it is the opposite of high-temperature long-day plants. This study systematically revealed the response expression characteristics of chia *SOC1* subfamily to four phytohormones under LDs and SDs.

This study showed that, in short-day plant chia, *ShSOC1-1* and *ShSOC1-2* are very different from each other in response to circadian rhythm and seasonal transition of the long-short photoperiod during the summer to autumn. *ShSOC1-1* is characterized by predominant expression at nights on sunny days but suppressed by rainy weather, while *ShSOC1-2* is inhibited by high temperatures sunny days in summer, and is dominantly expressed in rainy afternoon and in midnight of autumnal equinox (critical time for determining flowering and early floral differentiation). In SDs, short-day plant soybean *SOC1* and *SOC1*-like family genes were low during the daytime, high at night, peaked at midnight, and then decreased, with soybean *SOC1* consistent with chia *SOC1* genes while *SOC1*-like largely different from chia *SOC1* genes. In LDs, soybean *SOC1* and *SOC1*-like are dynamic and stable, with *SOC1*-like being similar to chia *SOC1* genes while *SOC1* being totally different from the chia *SOC1* trend of low-in-daytime and high-at-night (Na *et al.* 2013). Short-day plant *Zea mays ZmMADS1* as a homologue of *A. thaliana SOC1*, was low during the daytime and high at night (Alter *et al.* 2016), which consistent with chia *SOC1*, regardless of LDs or SDs.

At 37°C heat treatment, the expression of *ShSOC1-1* was firstly rapidly upregulated, but then it was greatly inhibited, while *ShSOC1-2* was immediately down-regulated and was successively inhibited. *A. thaliana SOC1* in long-day heat treatment was down-regulated, which was similar to chia (Takato *et al.* 2013). At 4°C cold treatment, *ShSOC1-1* was continuously upregulated, while *ShSOC1-2* was rapidly down-regulated, then increased continuously, and returned to basal level at 48 h, consistent with the trend of sustained up-regulation of *Arabidopsis SOC1* at 4°C cold treatment under LDs (Li *et al.* 2017). After PEG treatment, *ShSOC1-1* and *ShSOC1-2* had no major change and remained stable. After ABA and NaCl treatments, *ShSOC1-1* and *ShSOC1-2* were firstly down-regulated and even fluctuated in midway, but returned to near basal levels by 48 h, which was slightly different from that of *Arabidopsis*. *SOC1* was promoted in low and medium concentrations of NaCl, and inhibited in high concentrations of NaCl under LDs (Liu *et al.* 2013). After treatments with MeJA, SA, and mechanical injury, *ShSOC1-1* and *ShSOC1-2* was upregulated, especially when SA treatment was performed for 48 h, *ShSOC1-1* and *ShSOC1-2* was upregulated by 5–10 folds. This shows that *SOC1* is an integrator gene that

responds to the photoperiod, vernalization, autonomous, and gibberellin pathway, and its expression characteristics have certain fluctuations in the adverse conditions. *ShSOC1-1* is continuously upregulated in response to cold treatment, suggesting that it may also be affected by vernalization pathway. Inhibition of flowering genes are inhibited under low temperature conditions, thereby promoting expression of *SOC1* (Amasino 2005).

In this study, representative species with genome being completely sequenced in a typical taxonomic unit were selected (while a few representative species did not have genome sequencing, but the *SOC1* subfamily was cloned), and the phylogenetic tree of SOC1 proteins in the plant kingdom was constructed. The phylogenetic trees using the SOC1 proteins are in good agreement with the established plant community classifications in the academic community (Li *et al.* 2012; Zhong *et al.* 2012; Zhang *et al.* 2018). It also reveals some new features of *SOC1* gene evolution in the plant kingdom. There is no *SOC1* in non-seed plants (green aglae, moss, ferns, *etc*.), and all seed plants (gymnosperms, angiosperms) have *SOC1*. Because the angiosperms have a true floral structure, the "flowers" of gymnosperms are only spore-containers instead of a true flower structure, so the *SOC1* gene origin exactly accompanied with the origin of seed reproduction, and its function is to initiate sexual reproductive growth (strobile/flower differentiation) in gymnosperms and angiosperms in response to the best seasonal conditions, and creates conditions for follow-up reproductive behavior (strobile/flower opening, pollination, fertilization, and seed setting). Gymnosperm *Cryptomeria japonica SOC1*-like gene *CjMADS15* and *AGL6*-like gene *CjMADS14* regulate the development of male and female strobili (Katahata *et al.* 2014). Barley research also suggests that *SOC1* subfamily genes not only respond to vernalization and regulate flowering but also participate in regulating seed development (Papaefthimiou *et al.* 2012). The plants *SOC1* phylogenetic tree have two clusters: gymnosperms and angiosperms. The gymnosperm *SOC1* cluster occupies the basal site (Zhong *et al.* 2012), and the angiosperm *SOC1* cluster is divided into 4 branches. The monocots have 1 branch and the dicots have 3 branches corresponding to *Arabidopsis SOC1/AGL20*, *AGL42/71/72/SOC1*-like, and *AGL14/19/SOC1*-like, respectively. Within gymnosperms and monocots, it seems that *SOC1* duplication events widely occurred at order or family levels, resulting in 2 or more *SOC1* genes in most analyzed species. The ancestors of dicots have twice *SOC1* duplication events, making the 3-*SOC1*-gene basic status of all eudicots. The *AGL14/19* and *AGL42/71/72* duplication phenomena occurred in more than one families in the Brassicales order. However, only one orthologous gene corresponds to *AGL14/19* and *AGL42/71/72*, respectively, in non-Brassicales dicots. *AGL71/72* originated from *AGL42* duplication. At protein sequence level, AGL42 is conserved, while AGL71/72 diverged rapidly. It is noteworthy that the basal angiosperm *Amborella trichopoda* (common ancestor of monocots and dicots) has only one *SOC1* gene, indicating that angiosperms *SOC1* duplication events occurred after separation from basal angiosperms, and the non-duplicated and non-diversified status of important floral genes such as *SOC1* might be associated with the primitive and simple floral traits of basal angiosperms. The phylogenetic tree also showed that *ShSOC1-1* and *ShSOC1-2* were orthologous to *A. thaliana SOC1/AGL20* and *AGL42/71/72*, respectively. Since the orthologous gene corresponding to *AtAGL14/19* is common in dicots, it is speculated that this orthologous gene also exists in chia and needs to be cloned.

Two duplication events of *SOC1* in early dicots resulted in three *SOC1* genes in eudicots, and functional divergence occurred, *e.g.* some paralogs were also involved in regulating other traits besides flowering time. *Arabidopsis SOC1* subfamily controls flowering via regulating the expression of genes from the same family, *i.e. AGL42, AGL71, AGL72 etc*. (Dorca-Fornell *et al.* 2011), regulates leaf stomata opening, and prevents dark-induced chlorosis and senescence in leaves (Chen *et al.* 2017). *AGL14* regulates auxin polarity transport and root growth in addition to flowering (Garay-Arroyo *et al.* 2013). The 3 *P. hybrida SOC1*-like genes regulate flowering redundantly, but *FBP21* and *UNS* are related to developmental age, whereas *FBP28* is more related to short-day flowering habits (Preston *et al.* 2014). All of the 3 *Prunus mume SOC1*-like genes promoted flowering in transgenic *Arabidopsis*, but *PmSOC1-1* and *PmSOC1-2* also changed flower morphology, while *PmSOC1-3* did not (Li *et al.* 2016b). *Daucus carota DcSOC1-1* was related to early bolting and flowering with significant variations among different materials, while *DcSOC1-2* expression was low. In perennial short-day wild strawberry, *FvSOC1* regulates both vegetative and reproductive growth (Mouhu *et al.* 2013). The daily expression rhythm of *ParSOC1* in leaves of perennial trees *P. armeniaca* is related to cold demand and dormancy disruption (Trainin *et al*. 2013). *P. mume* SOC1 interacts with DAM6 to regulate both vegetative and floral bud differentiation (Kitamura *et al.* 2016). Divergence of expression patterns between *ShSOC1-1* and *ShSOC1-2* is similar to those observed in wild strawberry (Mouhu *et al.* 2013), Orchid *Dendrobium* (Ding *et al.* 2013) and *P. praecox* (Liu *et al.* 2016b). *ShSOC1-1* expression is high in stems and buds, considerable in mature leaves, semi-mature buds, mature buds and flowers, and low in seeds. *ShSOC1-2* expression is high only in young buds, but low in all other organs especially in seeds. Expression of *ShSOC1-1* in various organs is generally higher than that of *ShSOC1-2*, implying that *ShSOC1-1* is more important than *ShSOC1-2* in regulating whole-plant functions and plays a basic role in determining flowering. *ShSOC1-2* is sharply up-regulated in early stage of bud formation, which might play a decisive role as the "last straw" in flowering time, *i.e*. elevating the cellular *SOC1* level above a threshold in order to initiate flower bud differentiation.

Conclusion

In this study two *SOC1* genes, *ShSOC1-1* and *ShSOC1-2*, from chia were isolated and characterized. They had typical structural, molecular and expressional features, while distinct organ-specificity and responses to diverse physiological and environmental factors indicated their functional divergence. The effect of phytohormones on chia *SOC1* expression varied depending on the photoperiod. Chia *SOC1* expression also changed in response to circadian rhythms, climate and seasons, and was affected by a variety of abiotic stresses. This study also revealed some new evolutionary features, especially the origin and duplication, of plant-type *SOC1* genes. This study is the first report of *SOC1* subfamily genes of the order Lamiales. It will promote the flowering mechanism dissection of Lamiales, and also enrich the evolution and expression characteristics of plant *SOC1*. Our results will promote the study photoperiodic influence on flowering from the interaction between the photoperiod and the hormone pathways, and shed light on the molecular basis of flowering induction pathways in Chia and other short-day plants.

Acknowledgements

This study was supported by the Chongqing Research Program of Basic Research and Frontier Technology (cstc2015jcyjBX0143), National Key R&D Program of China (2016YFD0100506) and the Fundamental Research Funds for the Central Universities (No. XDJK2014D009).

References

Alter P, S Bircheneder, LZ Zhou, U Schlüter, M Gahrtz, U Sonnewald, T Dresselhaus (2016). Flowering time regulated genes in maize include the transcription factor *ZmMADS1*. *Plant Physiol* 172:389–404

Amasino RM (2005). Vernalization and flowering time. *Curr Opin Biotechnol* 16:154–158

Andersen CH, CS Jensen, K Petersen (2004). Similar genetic switch systems might integrate the floral inductive pathways in dicots and monocots. *Trends Plant Sci* 9:105–107

Ayerza R, W. Coates (2005). *Chia: Rediscovering a Forgotten Crop of the Aztecs*, pp: 1–197. The University of Arizona Press, Tucson, Arizona, USA

Bai S, JG Wallis, P Denolf, J Browse (2016). Directed evolution increases desaturation of a cyanobacterial fatty acid desaturase in eukaryotic expression systems. *Biotechnol Bioeng* 113:1522–1530

Chen J, X Zhu, J Ren, K Qiu, Z Li, Z Xie, J Gao, X Zhou, B Kuai (2017). *Suppressor of Overexpression of CO 1* negatively regulates dark-induced leaf degreening and senescence by directly repressing pheophytinase and other senescence-associated genes in *Arabidopsis*. *Plant Physiol* 173:1881–1891

Cortés D, HS Robledo, C Baginsky, L Morales-Salinas (2017). Climatic zoning of chia (*Salvia hispanica* L.) in Chile using a species distribution model. *Span J Agric Res* 15; Article e0302

Ding L, Y Wang, H Yu (2013). Overexpression of *DOSOC1*, an ortholog of *Arabidopsis SOC1*, promotes flowering in the orchid *Dendrobium* Chao Parya Smile. *Plant Cell Physiol* 54:595–608

Dorca-Fornell C, V Gregis, V Grandi, G Coupland, L Colombo, MM Kater (2011). The *Arabidopsis SOC1*-like genes *AGL42*, *AGL71* and *AGL72* promote flowering in the shoot apical and axillary meristems. *Plant J* 67:1006–1017

Garay-Arroyo A, E Ortiz-Moreno, MDLP Sánchez, AS Murphy, B García-Ponce, N Marsch-Martínez, SD Folter, A Corvera-Poiré, F Jaimes-Miranda, MA Pacheco-Escobedo, JG Dubrovsky, S Pelaz, ER Álvarez-Buylla (2013). The MADS transcription factor *XAL2/AGL14* modulates auxin transport during *Arabidopsis* root development by regulating *PIN* expression. *EMBO J* 32:2884–2895

Hepworth SR, F Valverde, D Ravenscroft, A Mouradov, G Coupland (2002). Antagonistic regulation of flowering-time gene *SOC1* by *CONSTANS* and *FLC* via separate promoter motifs. *EMBO J* 21:4327–4337

Hong JK, SY Kim, KS Kim, SJ Kwon, JS Kim, AK Jin, SI Lee, YH Lee (2013). Overexpression of a *Brassica rapa* MADS-box gene, *BrAGL20*, induces early flowering time phenotypes in *Brassica napus*. *Plant Biotechnol Rep* 7:231–237

Immink RGH, D Posé, S Ferrario, F Ott, K Kaufmann, FL Valentim, SD Folter, FVD Wal, ADJV Dijk, M Schmid, GC Angenent (2012). Characterization of *SOC1*'s central role in flowering by the identification of its upstream and downstream regulators. *Plant Physiol* 160:433–449

Jia HW, D Luo, B Cong, ZZ Gao, JY Shao, KM Cao, CR Sun (2000). Cloning and expression of a new MADS-box gene from rice. *Acta Bot Sin* 42:490–495

Katahata SI, N Futamura, T Igasaki, K Shinohara (2014). Functional analysis of *SOC1*-like and *AGL6*-like MADS-box genes of the gymnosperm *Cryptomeria japonica*. *Tree Genet Genomics* 10:317–327

Kim W, D Latrasse, C Servet, DX Zhou (2013). *Arabidopsis* histone deacetylase *HDA9* regulates flowering time through repression of *AGL19*. *Biochem Biophys Res Commun* 432:394–398

Kitamura Y, T Takeuchi, H Yamane, R Tao (2016). Simultaneous down-regulation of *DORMANCY-ASSOCIATED MADS-BOX6* and *SOC1* during dormancy release in japanese apricot (*Prunus mume*) flower buds. *J Hortic Sci Biotechnol* 91:476–482

Lee H, SS Suh, E Park, E Cho, JH Ahn, SG Kim, JS Lee, YM Kwon, I Lee (2000). The AGAMOUS-like 20 MADS domain protein integrates floral inductive pathways in *Arabidopsis*. *Gene Dev* 14:2366–2376

Li C, H Gu, W Jiang, C Zou, D Wei, Z Wang, Q Tang (2019). Protein interactions of SOC1 with SVP are regulated by a few crucial amino acids in flowering pathways of *Brassica juncea*. *Acta Physiol Plantarum* 41:43

Li H, K Ye, Y Shi, J Cheng, X Zhang, S Yang (2017). BZR1 positively regulates freezing tolerance *via* CBF-dependent and CBF-independent pathways in *Arabidopsis*. *Mol Plant* 10:545–559

Li W, X Liu, Y Lu (2016a). Transcriptome comparison reveals key candidate genes in response to vernalization of Oriental lily. *BMC Genomics* 17; Article 664

Li Y, Z Xu, W Yang, T Cheng, J Wang, Q Zhang (2016b). Isolation and functional characterization of *SOC1*-like genes in *Prunus mume*. *J Amer Soc Hortic Sci* 141:315–326

Li ZN, JG Zhang, GF Liu, XM Li, C Lu, JW Zhang, MZ Bao (2012). Phylogenetic and evolutionary analysis of A-, B-, C- and E-Class MADS-box genes in the basal eudicot *Platanus acerifolia*. *J Plant Res* 125:381–393

Liu C, C Zhu, HM Zeng (2016a). Key *KdSOC1* gene expression profiles during plantlet morphogenesis under hormone, photoperiod, and drought treatments. *Genet Mol Res* 15:1–24

Liu SN, TT Qi, JJ Ma, TF Ma, LY Ma, XC Lin (2016b). Ectopic expression of a *SOC1* homolog from *Phyllostachys violascens* alters flowering time and identity of floral organs in *Arabidopsis thaliana*. *Trees* 30:2203–2215

Liu T, Y Li, J Ren, Y Qian, X Yang, W Duan, X Hou (2013). Nitrate or NaCl regulates floral induction in *Arabidopsis thaliana*. *Biologia* 68:215–222

Liu XR, T Pan, WQ Liang, L Gao, XJ Wang, HQ Li, S Liang (2016c). Overexpression of an orchid (*Dendrobium nobile*) *SOC1/TM3*-like ortholog, *Dnagl19*, in *Arabidopsis* regulates *HOS1-FT* expression. *Front Plant Sci* 7; Article 00099

Ma G, G Ning, W Zhang, J Zhan, H Lv, M Bao (2011). Overexpression of *Petunia SOC1*-like gene *FBP21* in tobacco promotes flowering without decreasing flower or fruit quantity. *Plant Mol Biol Rep* 29:573–581

Moon J, SS Suh, H Lee, KR Choi, CB Hong, NC Paek, SG Kim, I Lee (2003). The *SOC1* MADS-box gene integrates vernalization and gibberellin signals for flowering in *Arabidopsis*. *Plant J* 35:613–623

Mouhu K, T Kurokura, EA Koskela, VA Albert, P Elomaa, T Hytönen (2013). The *Fragaria vesca* homolog of *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1* represses flowering and promotes vegetative growth. *Plant Cell* 25:3296–3310

Na X, B Jian, W Yao, C Wu, W Hou, B Jiang, Y Bi, T Han (2013). Cloning and functional analysis of the flowering gene *GmSOC1*-like, a putative *SUPPRESSOR OF OVEREXPRESSION CO1/AGAMOUS*-like *20* (*SOC1/AGL20*) ortholog in soybean. *Plant Cell Rep* 32:1219–1229

Papaefthimiou D, A Kapazogloua, AS Tsaftaris (2012). Cloning and characterization of *SOC1* homologs in barley (*Hordeum vulgare*) and their expression during seed development and in response to vernalization. *Physiol Plantarum* 146:71–85

Pearce S, LS Vanzetti, J Dubcovsky (2013). Exogenous gibberellins induce wheat spike development under short days only in the presence of *VERNALIZATION1*. *Plant Physiol* 163:1–13

Preston JC, SA Jorgensen, SG Jha (2014). Functional characterization of duplicated *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1*-Like genes in *Petunia*. *PLoS One* 9; Article e96108

Richter R, A Kinoshita, C Vincent, R Martinez-Gallegos, H Gao, ADV Driel, Y Hyun, JL Mateos, G Coupland (2019). Floral regulators *FLC* and *SOC1* directly regulate expression of the B3-type transcription factor *TARGET OF FLC AND SVP 1* at the *Arabidopsis* shoot apex via antagonistic chromatin modifications. *PLoS Genet* 15; Article e1008065

Richter R, E Bastakis, C Schwechheimer (2013). Cross-repressive interactions between *SOC1* and the *Gatas GNC* and *GNL/CGA1* in the control of greening, cold tolerance, and flowering time in *Arabidopsis*. *Plant Physiol* 162:1992–2004

Ruokolainen S, YP Ng, VA Albert, P Elomaa, TH Teeri (2011). Over-expression of the *Gerbera hybrida At-SOC1-like1* gene *Gh-SOC1* leads to floral organ identity deterioration. *Ann Bot* 107:1491–1499

Shi C, L Zhao, X Zhang, G Lv, Y Pan, F Chen (2019). Gene regulatory network and abundant genetic variation play critical roles in heading stage of polyploidy wheat. *BMC Plant Biol* 19; Article 6

Sreedhar RV, P Kumari, SD Rupwate, R Rajasekharan, M Srinivasan (2015). Exploring triacylglycerol biosynthetic pathway in developing seeds of chia (*Salvia hispanica* L.): A transcriptomic approach. *PLoS One* 10; Article 0123580

Takato H, M Shimidzu, Y Ashizawa, H Takei, S Suzuki (2013). An acyl-CoA-binding protein from grape that is induced through ER stress confers morphological changes and disease resistance in *Arabidopsis*. *J Plant Physiol* 170:591–600

Tan FC, SM Swain (2007). Functional characterization of *AP3*, *SOC1* and WUS homologues from citrus (*Citrus sinensis*). *Physiol Plantarum* 131:481–495

Tang Y, Y Jiang, J Meng, J Tao (2018). A brief review of physiological roles, plant resources, synthesis, purification and oxidative stability of Alpha-linolenic Acid. *Emir J Food Agric* 30:341–356

Trainin T, I Bar-Ya'Akov, D Holland (2013). *ParSOC1*, a MADS-box gene closely related to *Arabidopsis AGL20/SOC1*, is expressed in apricot leaves in a diurnal manner and is linked with chilling requirements for dormancy break. *Tree Genet Genomics* 9:753–766

Voogd C, T Wang, E Varkonyi-Gasic (2015). Functional and expression analyses of kiwifruit *SOC1*-like genes suggest that they may not have a role in the transition to flowering but may affect the duration of dormancy. *J Exp Bot* 66:4699–4710

Wei JY, DB Liu, GY Liu, J Tang, YY Chen (2016). Molecular cloning, characterization, and expression of *Misoc1*: A homolog of the flowering gene *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* from mango (*Mangifera indica* L). *Front Plant Sci* 7; Article 1758

Xue YF, N Yin, B Chen, F Liao, AN Win, J Jiang, R Wang, X Jin, N Lin, Y Chai (2017). Molecular cloning and expression analysis of two *FAD2* genes from chia (*Salvia hispanica*). *Acta Physiol Plantarum* 39; Article 95

Yoo SK, KS Chung, J Kim, JH Lee, SM Hong, SJ Yoo, SY Yoo, JS Lee, JH Ahn (2005). *CONSTANS* activates *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* through *FLOWERING LOCUS T* to promote flowering in *Arabidopsis*. *Plant Physiol* 139:770–778

Zhang X, J Wei, S Fan, M Song, C Pang, H Wei, C Wang, S Yu (2016). Functional characterization of *GhSOC1* and *GhMADS42* homologs from upland cotton (*Gossypium hirsutum* L.). *Plant Sci* 242:178–186

Zhang Y, D Tang, X Lin, M Ding, Z Tong (2018). Genome-wide identification of MADS-box family genes in moso bamboo (*Phyllostachys edulis*) and a functional analysis of *PeMADS5* in flowering. *BMC Plant Biol* 18; Article 176

Zhong X, X Dai, J Xv, H Wu, B Liu, H Li (2012). Cloning and expression analysis of *Gmgal1*, *SOC1* homolog gene in soybean. *Mol Biol Rep* 39:6967–6974