Seed Science Research (2007) **17**, 297–302 DOI: 10.1017/S0960258507836282

SHORT COMMUNICATION

Molecular cloning and characterization of a *LEAFY*-like gene highly expressed in developing soybean seeds

Qingchang Meng, Chunhong Zhang, Fang Huang, Junyi Gai and Deyue Yu*

National Center for Soybean Improvement, National Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, China

Abstract

LEAFY (LFY)-like proteins are plant-specific transcription factors that play essential roles in plant growth and development. In this study, a *LEAFY* homologue in soybean [*Glycine max* (L.) Merr.], designated as *GmLFY*, was cloned from inflorescences by the rapid amplification of cDNA ends (RACE) method. Sequence analysis showed that *GmLFY* cDNA contained a 1221 bp open reading frame, encoding 407 amino acid residues with typical characteristics of transcription factors. Subcellular localization in onion epidermal cells indicated that GmLFY protein was located in the nucleus. Analysis of the *GmLFY* genomic structure showed that the *GmLFY* gene has two introns with similar spliced sites as *LFY*-like genes in other plants. Reverse transcription-polymerase chain reaction (RT-PCR) analysis indicated that *GmLFY* was predominantly expressed in reproductive organs, such as inflorescences, pods and developing seeds. Further analysis showed that the expression level of *GmLFY* was higher in the middle stage than in the early or late stages of seed development, suggesting an essential regulatory role of GmLFY in soybean seed development.

Keywords: *Glycine max*, *LEAFY*, seed development, soybean, transcription factor

Introduction

Detailed analysis of completely sequenced genomes of model plants, such as *Arabidopsis* and rice, revealed that a significant number of the encoded proteins correspond to transcription factors (Riechmann and Ratcliffe, 2000). LEAFY/FLORICAULA (LFY/FLO) homologues, comprising a family of transcription factors unique to plants (Jack, 2004; Maizel *et al*., 2005), have been implicated in the regulation of gene expression in floral inductive pathways (Blázquez and Weigel, 2000), activation of downstream of floral homeotic genes (Weigel *et al*., 1992), cell–cell signalling and movement (Sessions *et al*., 2000; Wu *et al*., 2003), compound leaf development (Hofer *et al*., 1997; Dong *et al*., 2005), spikelet development (Kyozuka *et al*., 2000; Shitsukawa *et al*., 2006) and GA signalling (Blázquez and Weigel, 2000).

Although members of the LFY/FLO family have been isolated from many plants (Maizel *et al*., 2005), so far no information is available on *LFY* homologues in soybean, although this crop is considered as one of the main sources of protein and oil. As a first step in elucidating the function of LFY homologues in soybean, we characterized *GmLFY*, a soybean homologous *LFY* gene, and demonstrated that *GmLFY* was expressed mainly in reproductive organs, especially in seeds at the middle stage of development.

Materials and methods

Plant material

Soybean (*Glycine max* L.) cv. Bogao seedlings were grown in the field under normal conditions at the Nanjing Agricultural University. Vegetative tissues such as leaves, stems and roots were collected from 4-week-old seedlings, while young inflorescences at the R1 stage (Fehr *et al*., 1971), young pods at the R3 stage and developing seeds at 20 d after flowering (DAF) were collected and frozen in liquid nitrogen. Complete flowers at the R2 stage were separated into

*Correspondence
Fax: 86-25-84396410
Email: dyyu@njau.edu.cn

https://doi.org/10.1017/S0960258507836282 Published online by Cambridge University Press

four parts (sepals, petals, stamens and carpels). Seeds from different stages of development (15–50 DAF) were collected.

Isolation of genomic DNA and synthesis of cDNA

Genomic DNA was extracted from young leaflets of soybean using the CTAB (cetyl trimethyl ammonium bromide) protocol as described by Sambrook *et al.* (1989) with minor modifications. Total RNA was isolated using the RNeasy Plant Mini Kit (QIAGEN, Hilden, Germany), and subsequently treated with 10 units of RNase-free DNase I (TaKaRa, Otsu, Japan) per mg of RNA. First-strand cDNA was synthesized by incubation at 42 °C for 1 h in a final volume of 20 μl containing 4 μl 5 × buffer, 1 μg of total RNA, 500 μM oligo(dT)$_{18}$ primer, 10 units of AMV reverse transcriptase (TaKaRa), 1 mM dNTPs, and 20 units of RNase Inhibitor (Promega, Madison, Wisconsin, USA).

Sequence retrieval and primer design

GenBank databases were searched with the tBLASTn program (Altschul *et al.*, 1997) using both *Arabidopsis* LFY (Weigel *et al.*, 1992) and *Antirrhinum* FLO protein sequences (Coen *et al.*, 1990) as queries. One of the resulting soybean expressed sequence tags (ESTs), BU761377, was selected for primer design using the on-line program primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi), for rapid amplification of cDNA ends (RACE-PCR).

Cloning of GmLFY *cDNA*

In order to isolate full-length cDNA of *GmLFY*, both 5′ and 3′ RACE reactions were carried out according to the SMART RACE cDNA Amplification Kit User Manual (Clontech, Mountain view, California, USA). 5′ RACE reactions were performed to amplify the 5′ region of *GmLFY* with primers GSPR (5′-AGCTCGTTTGAGACCTCCTCGTCCA-3′) and NGSPR (5′-GCAGTGCACGTAGTGTCGCATCTTTG-3′). For 3′ RACE reactions, a gene-specific primer, GSPF (5′-GGCGAAGTTGCACGTGGAAAGAAG-3′) was used. The 5′- and 3′-RACE products were analysed on a 1.0% agarose/ethidium bromide gel, and the corresponding DNA band was purified, cloned into a pGEM-T vector (Promega) and transformed into the *Escherichia coli* strain DH5α. The positive clones were directly sequenced by Shanghai Invitrogen (Shanghai, China).

Subcellular localization of GmLFY

The putative open reading frame (ORF) of *GmLFY* was cloned into the pBI121 expression vector. The *GUS* gene was replaced with green fluorescent protein gene (*GFP*) to allow expression of GmLFY as a fusion protein with GFP under the control of the cauliflower mosaic virus 35S promoter. The generated fusion plasmid was used to transform onion epidermal cells, as described by Niu *et al.* (2005). Transient expression of GFP fusion was visualized by confocal laser scanning microscopy (Leica TCS SP2, Mannheim, Germany).

Semi-quantitative RT-PCR

Polymerase chain reactions (PCR) were conducted in a final volume of 20 μl using the transcribed templates. A pair of gene-specific primers: GmF (5′-GAGCAATGCCGTGAGTTCTT-3′) and GmR (5′-TTCTCTCAGCGTGACAAAGC-3′) were used for RT-PCR analysis. The soybean *actin* gene was selected as an internal control (Aeschbacher *et al.*, 1999). Amplification conditions were as follows: pre-denaturation at 94 °C for 5 min, followed by 35 cycles of 94 °C for 30 s, 53 °C for 30 s, and 72 °C for 1 min, with a final extension step of 72 °C for 10 min. PCR products were resolved on a 1.0% (w/v) agarose gel. The gel images were captured using a video camera (Image Master VDS, Amersham Pharmacia Biotech, Uppsala, Sweden). Quantification of the bands was performed using Quantity One software (Bio-Rad, Hercules, California, USA). The ratio between the signal intensities (with standard errors) of PCR product and of *actin* was calculated.

Results and discussion

Molecular cloning of a cDNA sequence homologous to the LFY/FLO *gene*

LFY/FLO-like proteins are among the most important transcription factors involved in floral development in plants. By searching the soybean EST database in GenBank with more than 350,000 soybean ESTs deposited (released on 19 April 2005), we found that only two accessions, BU761377 and CO979331, contained the putative *LFY/FLO*-like sequences of soybean, suggesting that the expression level of the *LFY/FLO* homologue(s) in soybean was low. Comparison of these two sequences showed that they were identical. Therefore, the longer EST, BU761377, was used to design gene-specific primers for 5′ or 3′ RACE PCR reactions.

The longest cDNA, with 1515 base pairs (bp), was selected after RACE-PCR. It contained a 1221 bp open reading frame. The 5′ untranslated region of the cDNA sequence was 34 bp long upstream from the initiation ATG start codon located at position 35–37. After stop codon TAA, it was 257 bp long in the 3′ untranslated region, with 26 bp poly(A) flanking the 3′ end (data not shown). After the database search, it showed extensive similarity to previously published

https://doi.org/10.1017/S0960258507836282 Published online by Cambridge University Press

LFY/FLO-like genes and hence was tentatively designated as *GmLFY* (*Glycine max LFY* homologue).

The sequences near the translation start site correspond to the eukaryotic consensus sequence GCC(AG)CCATGG (Kozak, 1999). We selected three different long-fragment insertion clones of the 5′ RACE-PCR product for sequencing, and the results of sequencing showed that their sequence length before the putative translation start site in this study was very short because of a high GC content (data not

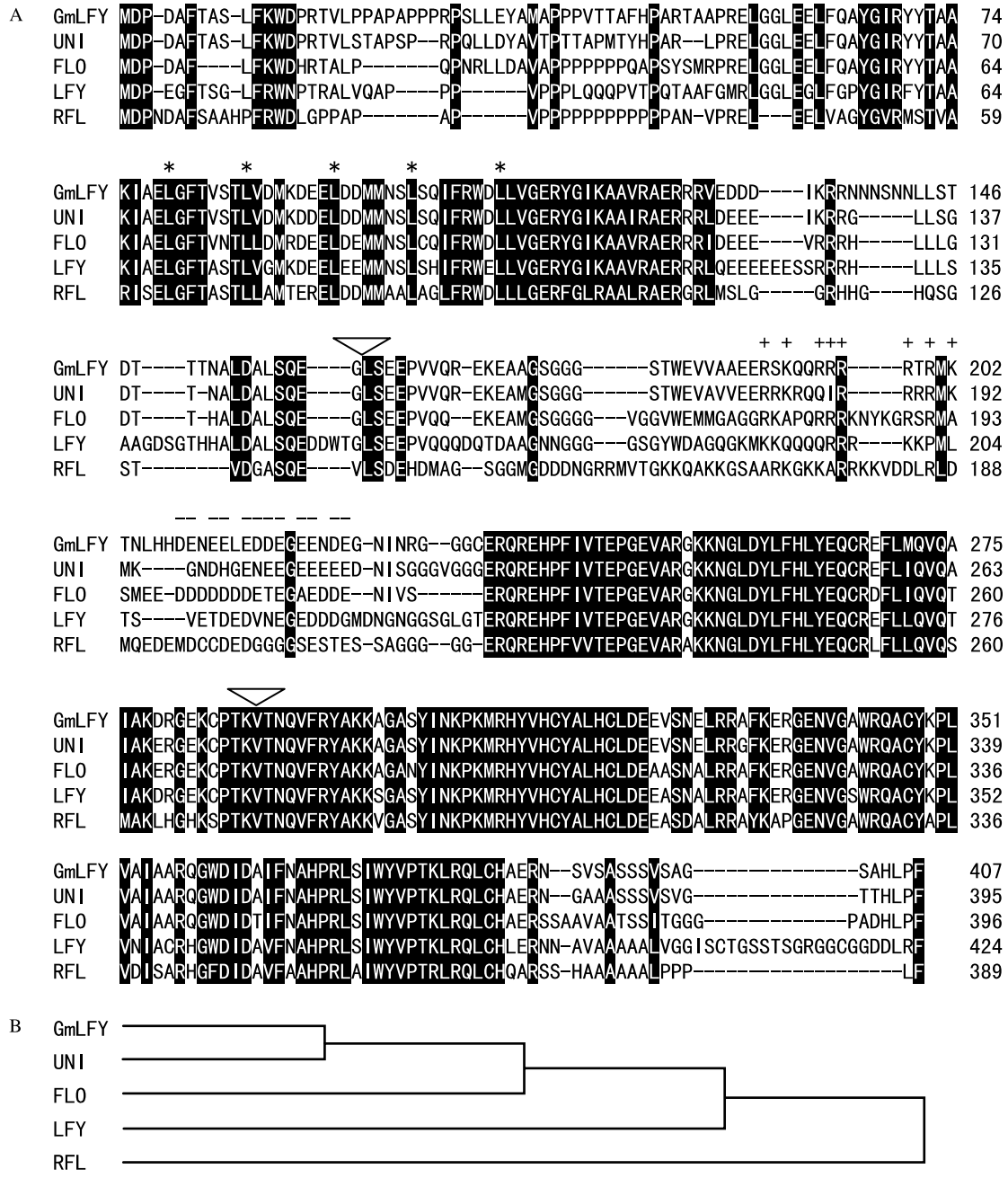


Figure 1. Alignment and dendrogram of several LFY/FLO proteins. (A) Comparison of the amino acid sequence of GmLFY with other LFY/FLO-like proteins: UNI in *Pisum sativum* (accession number: AF010190), FLO in *Antirrhinum majus* (accession number: M55525), LFY in *Arabidopsis thaliana* (accession number: M91208), and RFL in *Oryza sativa* (accession number: AB005620). The leucine repeats are marked with an asterisk. The basic and acidic regions are labelled ' + ' and '-', respectively. The conserved sites of introns are indicated by triangles. (B) Dendrogram of LFY/FLO proteins.

https://doi.org/10.1017/S0960258507836282 Published online by Cambridge University Press

shown). Furthermore, the sequences of LEAFY proteins are quite conserved (Fig. 1). The sequence length and homology of the translated product of *GmLFY* lead us to conclude that we have obtained a full-length clone.

Comparison of genomic and cDNA sequences established the exon/intron structure of the soybean *GmLFY* gene. The first intron was 483 bp in length, while the second was relatively longer (887 bp in length). Two introns contained the canonical GT–AG binucleotide splice site junctions (data not shown). The conserved splicing sites of introns were observed between *GmLFY* sequence and other published *LFY/FLO* homologues (Fig. 1A) after alignment. The genomic and cDNA sequences for *GmLFY* have been deposited in GenBank under the accession numbers DQ448809 and DQ448810, respectively.

The deduced GmLFY protein containing 407 amino acid residues had a molecular mass of 46.0 kDa. A search in GenBank protein databases revealed that GmLFY shared 73.6% identity to *Antirrhinum majus* FLO, and 63.7% to *Arabidopsis thaliana* LFY. High sequence similarity was obviously observed among the homologues in other legume species, such as *Medicago truncatula*, *Pisum sativum* and *Lotus japonicus*. It exhibited 83.9% sequence identity to UNI of *Pisum sativum*. Further comparison of GmLFY with data from conserved domain databases (CDD) revealed that GmLFY showed 69.8% identity to the 393 residue conserved domain of the LFY/FLO protein (CDD accession number 41731, data not shown) at its C-terminus. The high similarity between GmLFY and other LFY/FLO-like proteins suggests that they may have conserved functions. However, it showed relatively low sequence identity (55.1%) to RFL, a LFY/FLO homologue of rice. A phylogenetic tree was constructed using GmLFY protein and selected LFY/FLO proteins from other plants (Fig. 1B).

GmLFY showed the presence of several typical motifs characteristic of transcription factors (Fig. 1A). First, the GmLFY protein had a Pro-rich tract (20–29) in the N-terminus, which could be involved in transcriptional regulation (Esumi *et al.*, 2005). Secondly, a short leucine zipper of $Leu^{79, 86, 94, 102}$, followed by a basic region, essential for protein dimerization and DNA binding, was also present. Thirdly, the protein contained a highly acidic region (between Asp^{208} and Glu^{213}), a typical transcription activation domain, which appeared to interact with parts of the transcription complex (Weaver and Hedrick, 1997). Finally, the C-terminal region of GmLFY contained a serine-rich region (391–403), a putative activation domain. These motifs have been suggested to be required for the correct recognition and/or activation of promoter elements of downstream genes (Mitchell and Tjian, 1989).

Subcellular localization of the GmLFY–GFP fusion

LFY was identified as a transcription factor by *in vivo* analysis (Busch *et al.*, 1999; Wagner *et al.*, 1999). However, we failed to identify a nuclear localization signal (NLS) for GmLFY (data not shown). Therefore, to assess the subcellular localization of the GmLFY protein, a construct containing a translational fusion of its ORF to the GFP reporter gene under the control of a 35S promoter was used to transform onion epidermal layers. Confocal imaging of GFP fluorescence showed that the GmLFY–GFP fusion protein was located in the nucleus (Fig. 2). In contrast, the GFP fluorescence was distributed throughout the cells when the onion epidermal layers were treated with the 35S::GFP construct (Fig. 2A). This suggests that GmLFY was located in the nucleus, although it did not contain a typical plant NLS.

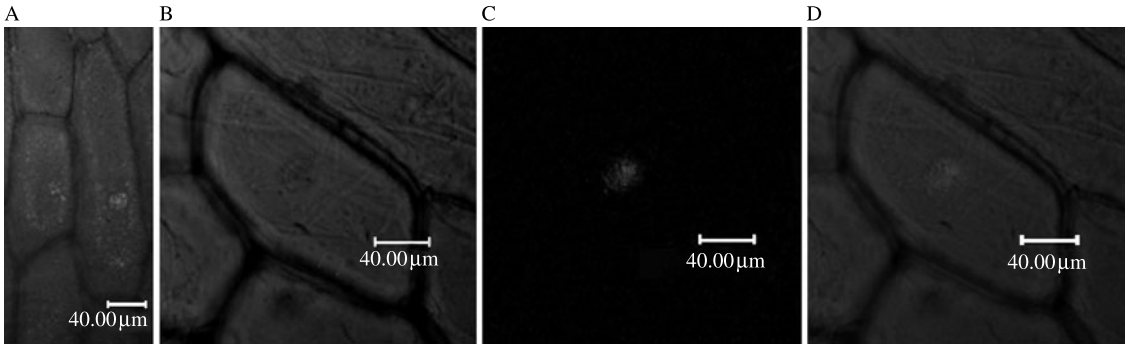


Figure 2. Subcellular localization of GmLFY protein. Onion epidermal cells were transiently transformed with 35S::GmLFY-GFP (panels B, C and D) and 35S::GFP (panel A). After incubation for 24 h, cells were observed with a confocal microscope. Panel B is a bright-field image; panel C is a fluorescence image, and panel D shows the merged images.

https://doi.org/10.1017/S0960258507836282 Published online by Cambridge University Press

Expression pattern of GmLFY *in soybean tissues*

A semi-quantitative RT-PCR method was used for quantification of *GmLFY* expression, with the *actin* gene as an internal control. For RT-PCR, we designed a pair of gene-specific primers that flanked the second intron to avoid genomic DNA contamination. As shown in Fig. 3, expression of *GmLFY* was clearly detected in all of the examined reproductive tissues, such as inflorescence (In), pods (Pd) and developing seeds (Sd). Furthermore, we detected the transcripts in stems, leaves and shoot tips at relatively low levels, whereas no *GmLFY* transcript was detectable in roots. These results indicate that *GmLFY* may play multiple roles in soybean development, especially in the reproductive tissues. Moreover, *GmLFY* was highly expressed in sepals and carpels, but showed very low abundance in petals and stamens (Fig. 3), which was very similar to the expression pattern of *ALF2*, a *LFY/FLO* orthologue of apple (Wada *et al.*, 2002). This indicates that *GmLFY*, as well as *ALF2*, were possibly required during late development of sepals, petals and carpels.

Expression of GmLFY *in developing soybean seeds*

We investigated the expression pattern of *GmLFY* in developing seeds (Fig. 4). The expression of *GmLFY* roughly followed a bell-shaped pattern over the developmental period. It was expressed at a medium level in seeds at 15 DAF and then decreased in 20 DAF seeds. Thereafter, it increased sharply, reaching its maximum between 35 and 40 DAF, and decreased immediately after that time point. The expression pattern and expression level of *GmLFY* in the developing seeds were apparently dependent on the status of seed growth during seed filling.

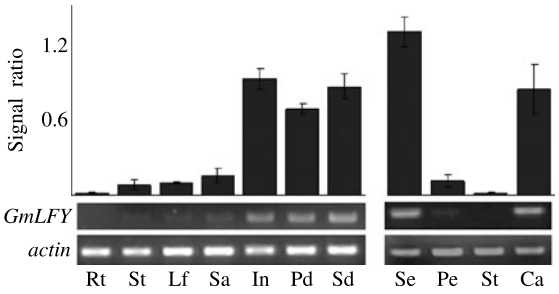


Figure 3. Tissue-specific expression of the *GmLFY* gene in soybean tissues. From the left: Rt, roots of seedlings; St, young stems; Lf, leaves; Sa, vegetative shoot apices; In, inflorescence apices; Pd, young pods without seeds; Sd, developing seeds at 15 DAF; Se, sepals; Pe, petals; St, stamens; Ca, carpels. Three replicates of each experiment were performed independently, but only one of the similar images is shown here. Standard errors are indicated.

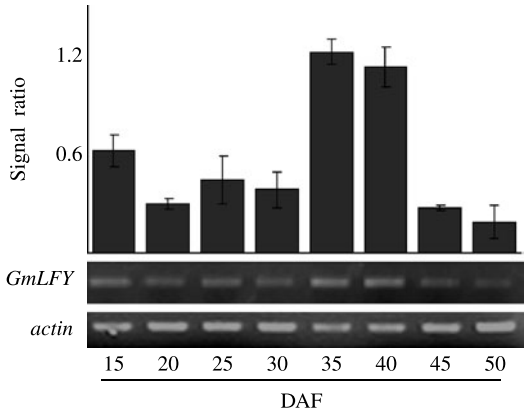


Figure 4. Expression of *GmLFY* in developing soybean seeds. Three replicates were independently analysed with similar results, and a representative experiment is shown here. Standard errors are indicated. DAF, days after flowering.

Although many *LFY/FLO*-like genes have been characterized, so far there are no reports of *LFY/FLO*-like genes expressed in seeds, using comprehensive genomic or proteomic approaches. Beilinson *et al.* (2005) identified a gene, differentially expressed in developing soybean seeds, which showed extensive similarity with the *TERMINAL FLOWER 1* (*TFL1*) gene in *Arabidopsis*. *TFL1* affects the developmental phases of *Arabidopsis* by delaying the up-regulation of *LFY* (Ratcliffe *et al.*, 1999). Although the *TFL*-like gene was expressed in soybean seeds at 21 DAF, a *LFY/FLO*-like gene had not yet been demonstrated to be expressed in seeds at this developmental stage. We found the transcripts of *GmLFY* in seeds, supporting the putative role of the *TFL1*-like gene in the regulation of *GmLFY* transcript abundance in soybean seeds. Further analysis should be carried out to dissect the roles of the *GmLFY* and *TFL1*-like genes in developing soybean seeds.

Acknowledgements

This work was supported in part by National 973 project (2004CB117206), 863 project (2006AA1021C1, 2006AA10A111), National Natural Science Foundation of China (30490250), and an award grant for Outstanding Scholars from the Ministry of Education of China.

References

Aeschbacher, R.A., Müller, J., Boller, T. and Wiemken, A. (1999) Purification of the trehalase GMTRE1 from soybean nodules and cloning of its cDNA. GMTRE1 is expressed at a low level in multiple tissues. *Plant Physiology* **119**, 489–496.

https://doi.org/10.1017/S0960258507836282 Published online by Cambridge University Press

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402.

Beilinson, V., Moskalenko, O.V., Ritchie, R.D. and Nielsen, N.C. (2005) Differentially expressed genes during seed development in soybean. *Physiologia Plantarum* **123**, 321–330.

Blázquez, M.A. and Weigel, D. (2000) Integration of floral inductive signals in *Arabidopsis*. *Nature* **404**, 889–892.

Busch, M.A., Bomblies, K. and Weigel, D. (1999) Activation of a floral homeotic gene in *Arabidopsis*. *Science* **285**, 585–587.

Coen, E.S., Romero, J.M., Doyle, S., Elliot, R., Murphy, G. and Carpenter, R. (1990) *floricaula*: a homeotic gene required for flower development in *Antirrhinum majus*. *Cell* **63**, 1311–1322.

Dong, Z.C., Zhao, Z., Liu, C., Luo, J., Yang, J., Huang, W., Hu, X., Wang, T.L. and Luo, D. (2005) Floral patterning in *Lotus japonicus*. *Plant Physiology* **137**, 1272–1282.

Esumi, T., Tao, R. and Yonemori, K. (2005) Isolation of *LEAFY* and *TERMINAL FLOWER 1* homologues from six fruit tree species in the subfamily Maloideae of the Rosaceae. *Sexual Plant Reproduction* **17**, 277–287.

Fehr, W.R., Caviness, C.E., Burmood, D.T. and Pennington, J.S. (1971) Stage of development descriptions for soybeans, *Glycine max* (L.) Merrill. *Crop Science* **11**, 929–931.

Hofer, J., Turner, L., Hellens, R., Ambrose, M., Matthews, P., Michael, A. and Ellis, N. (1997) *UNIFOLIATA* regulates leaf and flower morphogenesis in pea. *Current Biology* **7**, 581–587.

Jack, T. (2004) Molecular and genetic mechanisms of floral control. *Plant Cell* **16**, S1–S7.

Kozak, M. (1999) Initiation of translation in prokaryotes and eukaryotes. *Gene* **234**, 187–208.

Kyozuka, J., Kobayashi, T., Morita, M. and Shimamoto, K. (2000) Spatially and temporally regulated expression of rice MADS box genes with similarity to *Arabidopsis* class A, B, and C genes. *Plant and Cell Physiology* **41**, 710–718.

Maizel, A., Busch, M.A., Tanahashi, T., Perkovic, J., Kato, M., Hasebe, M. and Weigel, D. (2005) The floral regulator LEAFY evolves by substitutions in the DNA binding domain. *Science* **308**, 260–263.

Mitchell, P.J. and Tjian, R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**, 371–378.

Niu, X.G., Chen, Q.J., Liu, Q. and Wang, X.C. (2005) Cloning, subcellular localization and expression analysis of an inositol 1,3,4-trisphosphate 5/6-kinase-like gene in *Oryza sativa*. *Journal of Agricultural Biotechnology* **13**, 815–816 (in Chinese with English abstract).

Ratcliffe, O.J., Bradley, D.J. and Coen, E.S. (1999) Separation of shoot and floral identity in *Arabidopsis*. *Development* **126**, 1109–1120.

Riechmann, J.L. and Ratcliffe, O.J. (2000) A genomic perspective on plant transcription factors. *Current Opinion in Plant Biology* **3**, 423–434.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular cloning. A laboratory manual* (2nd edition). Plainview, NY, Cold Spring Harbor Laboratory Press.

Sessions, A., Yanofsky, M.F. and Weigel, D. (2000) Cell–cell signaling and movement by the floral transcription factors LEAFY and APETALA1. *Science* **289**, 779–782.

Shitsukawa, N., Takagishi, A., Ikari, C., Takumi, S. and Murai, K. (2006) *WFL*, a wheat *FLORICAULA/LEAFY* ortholog, is associated with spikelet formation as lateral branch of the inflorescence meristem. *Genes and Genetic Systems* **81**, 13–20.

Wada, M., Cao, Q., Kotoda, N., Soejima, J. and Masuda, T. (2002) Apple has two orthologues of *FLORICAULA/LEAFY* involved in flowering. *Plant Molecular Biology* **49**, 567–577.

Wagner, D., Sablowski, R.W.M. and Meyerowitz, E.M. (1999) Transcriptional activation of *APETALA1* by LEAFY. *Science* **285**, 582–584.

Weaver, R.F. and Hedrick, P.W. (1997) *Genetics* (3rd edition). Dubuque, Iowa, William C. Brown Publishing.

Weigel, D., Alvarez, J., Smyth, D.R., Yanofsky, M.F. and Meyerowitz, E.M. (1992) *LEAFY* controls floral meristem identity in *Arabidopsis*. *Cell* **69**, 843–859.

Wu, X.L., Dinneny, J.R., Crawford, K.M., Rhee, Y., Citovsky, V., Zambryski, P.C. and Weigel, D. (2003) Modes of intercellular transcription factor movement in the *Arabidopsis* apex. *Development* **130**, 3735–3745.

Received 5 March 2007
accepted after revision 9 August 2007
© 2007 Cambridge University Press

https://doi.org/10.1017/S0960258507836282 Published online by Cambridge University Press