Origin of *Chrysanthemum* cultivars — Evidence from nuclear low-copy *LFY* gene sequences

Abstract :

The chrysanthemums (*Chrysanthemum* × *morifolium* Ramat.) are a well-known group of traditional ornamental flowers in China and have been cultivated all over the world. Yet the origin of chrysanthemum cultivars has been highly debated. In this study we employ the nuclear low-copy *LFY* gene to study the evolutionary history of chrysanthemum cultivars. The structure of the *LFY* gene in all *Chrysanthemum* species examined is highly conserved with three exons and two introns. The length of the *LFY* gene in *Chrysanthemum* varied from 2, 887 to 3, 348 bp. The two introns exhibited high levels of variation in length and sequence composition at the intraspecific and interspecific levels. Phylogenetic analysis of the whole *LFY* sequences of *Chrysanthemum* resulted in topologies that contained three major clades. The *LFY* sequences from the same cultivars are present in two or three clades, supporting that hybridization and allopolyploidy were important mechanisms in the origins of different chrysanthemums. Our results suggest that different cultivars had different ancestors. *Chrysanthemum indicum*, *C. zawadskii* and *C. nankingense* were likely the direct ancestors of most chrysanthemum cultivars examined. *Chrysanthemum vestitum* is a putative ancestor for some cultivars, and may have indirectly involved in the development of the chrysanthemum cultivars. Sequences of the *LFY* gene are informative to shed insights into the origin of chrysanthemum

cultivars and show great potential as a phylogenetic marker to decipher the phylogeny of *Chrysanthemum* and its close relatives.

Key words: *Chrysanthemum* cultivars; origin; allopolyploidy; *LFY* gene

1. Introduction

As one of the ten most famous Chinese traditional flowers, *Chrysanthemum* × *morifolium* Ramat. is composed of many cultivars among the horticultural plants in the world (Chen 1985). It has been proposed that chrysanthemum is a hybrid cultigen complex derived from interspecific hybridization and artificial selection (Dai et al. 2002). With the long history of cultivation, directional selection, and genetic drift of a small group, the domestication process has resulted in the situation that many morphological characteristics of chrysanthemums differ substantially from those of the wild parental species (Dai et al. 2005; Ren and Guo 2015). The drastic morphological changes in the domestication have made it difficult to tease apart the origin of the Chinese chrysanthemums. Many studies including comparative morphology, cytology, interspecific hybridization, and molecular systematics have suggested that the hexaploid *Chrysanthemum vestitum* (Hemsl.) Ling et Shih and the tetraploid *C. indicum* (L.) Des Moul. may be major donors of the chrysanthemum cultivars; and other closely related wild species such as *C. nankingense*, *C. lavandulifolium* and *C. zawadskii* may also have involved in the domestication of the cultivated chrysanthemums (Fukai et al. 2002; Dai et al. 2005). The origin of the cultivated chrysanthemums remains unclear and needs to be explored with single-copy nuclear markers. In addition, there are various classifications on chrysanthemum cultivars based on flower color, type, size, flowering time and plant height, but they differ considerably without a generally accepted classification system for the chrysanthemum cultivars. Therefore, it is critical to find genetic markers for the identification of varieties and for the breeding of chrysanthemums.

The *LEAFY* (*LFY*) gene was first found in *Antirrhinum majus* L. and was designated to be *FLORICAULA* (Coen et al. 1990; Zimmer and Wen 2013). Then in *Arabidopsis thaliana* (L.) Heynh, it was named as *LEAFY* and was characterized as a floral meristem gene (Schultz and Haughn 1991). It encodes a plant-specific transcription factor that plays a key role in floral development (Weigel et al. 1992). The *LFY* gene regulates the establishment of the identity of the floral meristem as well as flowering time in *Arabidopsis* (Blazquez et al. 1997), and is active in organ identity (Blazquez and Weigel 2000; Winter et al. 2011). *LFY* genes have been identified in different plant families (Frohlich and Parker 2000; Rudall et al. 2009). The structure of *LFY* gene in angiosperms is well conserved with three exons and two introns (Frohlich and Meyerowitz 1997; Peng et al. 2010; Ma et al. 2014). It is single-copy in most diploid angiosperms (Ma et al. 2014). The size of intron 2 from 12 species of angiosperms has been surveyed and was found that the intron had a high nucleotide substitution rate (Frohlich and Meyerowitz 1997; Peng et al. 2010; Ma et al. 2014). The low-copy number and the relatively high nucleotide substitution rate in the introns of the gene in angiosperms made the introns of this gene a good candidate marker for phylogenetic reconstructions of closely related species, and they have been used for phylogenetic reconstruction in many plant groups (Tu et al. 2008; Peng et al. 2010; Zheng et al. 2011). In two diploid species of *C. nankingense* and *C. lavandulifolium*, *LFY* was present as a single-copy (Ma et al. 2008; Ma et al. 2013a). However, in the chrysanthemum cultivar ‘qianshouguanyin’, there were two copies and the two introns of *LFY* showed distinct variations (Ma et al. 2013b).

In this study, we cloned and analyzed the *LFY* homologue from several wild species of *Chrysanthemum* and some cultivars. Our primary objective is to address questions concerning the parental donors of chrysanthemum cultivars by comparing the sequences with those of the wild relatives. With several species sequenced for the first time, we also intend to evaluate the phylogenetic utility of *LFY* in *Chrysanthemum*.

2. MATERIAL AND METHODS

2.2 Plant material

Seven putative wild donor species, including the diploid species *Chrysanthemum lavandulifolium* and *C. nankingense*; tetraploid species *C. indicum* and *C. chanetii*; and hexaploid species *C. vestitum* and *C. zawadskii*, were collected from their native habitats. The cultivars in the present study were some of the main types of chrysanthemums and were hexaploids or hexaploid-based aneuploids. All plants were cultivated at the Flower Garden of Northeastern University, Shenyang, Liaoning Province, China (Table 1).

2.3 DNA extraction

Genomic DNA was extracted from young leaves using the modified CTAB method (Ma and Dai 2009). DNA quality and quantity were measured based on its absorbance at 260 and 280 nm with a Helena Biosciences Biophotometer (Eppendorf).

2.4 Amplification and sequencing

The *LFY* gene was initially amplified in two fragments. The primers LK and GSP2 were used to amplify the 5' fragment of *LFY*; and primers D1 and LB were used to

amplify the 3' fragment in *Chrysanthemum*. It was problematic to assemble the whole sequence from the two fragments. We thus then used long PCR to amplify the whole transcription of the *LFY* homologue from the different *Chrysanthemum* species, using the primer pair LK and LB. PCR reactions were conducted in a 50 µl volume with 5 µl Taq polymerase buffer, 100 ng total genomic DNA, 0.4 mM of both forward and reverse primers, 2 µl dNTPs (10mM, each dNTP), and two units of LA Taq DNA polymerase (Takara, Japan), with the program of an initial denaturation 5 min at 94°C, and then 30 cycles of denaturation 30 s at 94°C, annealing at 55°C for 30 s, and extension 3 min at 72°C. The PCR products were separated and cloned following the TOPO-XL-PCR Cloning Kit (Invitrogen, USA) protocol. Primers used for amplifying the *LFY* genes from different species in *Chrysanthemum* were presented in Table 2. Two pairs of nested primers 110 F and 110 R, D1 and GSP0 were used as internal primers to check the specificity of the amplicons. Transformed colonies were used by PCR with specific amplification and nested internal primers. Restriction digestion was also used to check the presence of inserts of the appropriate size. Both strands of the templates were sequenced by Zhongmeitaihe Gene Company (Beijing, China). The number of clones sequenced per species is provided in Table 1. Nucleotide sequences of different *LFY* haplotypes were submitted to GenBank at NCBI (National Center for Biological Information, Bethesda, MD, USA), and the GenBank accession number for each is listed in Table 1.

2.5 Phylogenetic analysis

The program 'DNAMAN' version 8.0 (Lynnon Biosoft Company, Canada) was used

to verify the 63 *LFY* sequences obtained from the *Chrysanthemum* species. BLAST searches were used to confirm if the obtained sequences were homologs of the *LFY* gene in GenBank (Altschul et al. 1990). The exon and intron regions were determined by comparing the sequences of the *LFY* gene in *C. nankingense* and *C. lavandulifolium* published earlier (Ma et al. 2008, 2013a). Sequences were aligned in the program Geneious v.8.1.2 (<http://www.geneious.com/> Kearse et al. 2012) using MUSCLE (Edgar 2004) and adjusted manually. Ambiguously aligned regions in the alignment were removed from the matrix prior to the phylogenetic reconstructions. PAUP* 4.0 (Swofford 2002) was used to calculate the numbers of variable and parsimony-informative sites. Phylogenetic analyses were conducted using Bayesian inference (BI) with MrBayes version 3.2.5 (Ronquist and Huelsenbeck 2003; Ronquist et al. 2012) and maximum likelihood (ML) with RAxML version 8.2 (Stamatakis 2014). For Bayesian inference, the dataset was partitioned into five parts, i.e., the three exons and two introns. The best-fit nucleotide substitution models for each of the five partitions were determined using the program jModelTest v. 2.1.7 by corrected Akaike Information Criterion (AICc) (Darriba et al. 2012). Model parameters were unlinked across the five partitions. Four separate Markov chain Monte Carlo (MCMC) runs were initiated, each with 10 000 000 generations. Runs were started from a random tree; the topology was sampled every 1000 generations of the MCMC chain. A majority-rule (50%) consensus tree and the posterior probabilities (PP) were constructed after removing the first 25% of the sampled trees as burn-in. The ML analysis was conducted in RAxML by applying the GTR+G

model to each partition of the dataset. The rapid bootstrap analysis was conducted with a random seed and 1000 alternative runs. *Chrysanthemum chanetii* was used to root the trees. Tree rooting and visualization were achieved in Fig Tree v1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>). Bootstrap support (BS) values from the ML analyses were labeled on the corresponding branches of the BI tree.

3. Results

3.1 Characteristics of *LFY* sequences in *Chrysanthemum* species

In the polyploids, more than one sequence type was present in the two introns. All the sequences contained three exons and two introns with positions conserved just as in the *LFY* gene of *C. nankingense* and *C. lavandulifolium* (Fig. 1). The sequences of the open reading frame had either 1236 bp or 1239 bp that encoded a putative protein of 412 or 413 amino acids and were more than 96% identical to each other in *Chrysanthemum*. All the putative proteins of *LFY* in different species of *Chrysanthemum* share specific regions such as putative short Leu zipper, basic region formed by a core of Arg and Lys residues and acidic region, which were also found in other angiosperm FLO/*LFY* proteins (data not shown). These data thus indicate that we isolated the *LFY* homolog from surveyed species of *Chrysanthemum* successfully. The length of the genomic sequences of *LFY* in *Chrysanthemum* ranged from 2, 887 bp to 3, 348 bp. The length and sequence composition of the three exons were conserved, with the exception of one of the sequences of ‘fenhuangxiaori’ that lacked eleven codons. However, both introns exhibited high levels of variation in length and sequence composition. The length of intron 1 ranged from 490 to 885 in various

species of *Chrysanthemum* and that of intron 2 ranged from 1066 to 1291 bp. There are multiple sections of indels in the aligned sequences. The structure scheme of the *LFY* gene in the genus *Chrysanthemum* is presented in Fig. 1 (modified from Ma et al. 2014).

3.2 Phylogenetic analysis of *LFY* sequences in *Chrysanthemum*

The *LFY* sequences of *Chrysanthemum*, including different clones from all the sampled individuals, were subjected to phylogenetic analysis. The best-fit nucleotide substitution models for each of the three exons and two introns were shown in Table S1. Based on these models, BI and ML analyses showed congruent results that yielded identical trees (Fig. 2). For the discussions below, we only present results obtained from the BI tree. Three distinct clades can be discerned in the BI tree. The chrysanthemum cultivars were found in three clades with different wild species. Clade I consists of tetraploid *C. indicum*, diploid *C. nankingense* and *C. lavandulifolium* and five of the six hexaploid chrysanthemum cultivars. Furthermore, the diploid *C. lavandulifolium* and three sequences of the tetraploid *C. indicum* formed a smaller subclade with 100% bootstrap support values. In addition, *C. nankingense* grouped with ‘fenzhouqiu’ and a ‘xiaoju’ with a high support value. The hexaploid *C. vestitum* and *C. argyrophyllum* grouped with chrysanthemum cultivars ‘fenzichiguan’, ‘fenzhouqiu’, and ‘fenhuanxiaori’, forming clade II. *Chrysanthemum zawadskii* and some chrysanthemum cultivars including *C. x morifolium* ‘fenzichiguan’ and ‘fenhuangxiaori’ formed clade III with 100% bootstrap support values.

4. Discussion

4.1 *LFY* gene structure and sequence variation in *Chrysanthemum*

All the *LFY* homologs isolated from species of *Chrysanthemum* based on this study as well as our previous work (Ma et al. 2008, 2013a, 2014) contained three exons and two introns. This is congruent with *LFY* homologs from other angiosperms, supporting that the *LFY* gene structure is conserved. *LFY* has been reported as a single-copy gene in most diploid angiosperms studied to date (Frohlich and Parker 2000; Ma et al. 2014; Zheng et al. 2011). Nevertheless, there is variation detected in the length and sequence composition of the *LFY* homologs across *Chrysanthemum* species. There were 641 variable characters, of which 397 were parsimony-informative characters in the aligned matrix of 63 sequences, largely as a result of differences within the two introns, which were highly variable (Table 3). Many indels and SNPs could be found in these two introns. It has been suggested that single-copy genes on homologous chromosomes of allopolyploids may be distinct (Liu et al. 2007; Ma et al. 2014). Predicted *LFY* proteins from the different species sampled in the study had more than 96% identity with each other, suggesting that the *LFY* homologs in different species of *Chrysanthemum* may have the same function as presented by Ma et al. (2014)

4.2 Origin of the cultivated chrysanthemums

Many studies have investigated the origin of chrysanthemums and it has been proposed that *C. vestitum* and *C. indicum* may be the major donors of chrysanthemum cultivars (Dai and Chen 1996; Dai et al. 1998; Dai et al. 2005). The phylogenetic analyses of the nuclear *LFY* sequences suggest that *C. indicum* was an ancestor which

contributed to the evolution of chrysanthemum cultivars ‘fenhuangxiaori’, ‘fenzhouqiu’, ‘guandongdaxia’, and ‘yulupantao’, and *C. nankingense* may have involved in the origin of ‘fenzhouqiu’ and a ‘xiaoju’. The result is consistent with previous investigations (Dai et al. 2005; Fukai et al. 2002). The *LFY* data further suggest that *C. zawadskii* is a likely ancestor of the cultivar chrysanthemums (Fig. 2). Although previous results proposed that *C. zawadskii* was a wild species remotely related to the modern cultivars of chrysanthemums (Dai and Chen 1996; Dai et al. 2005), *Chrysanthemum zawadskii* formed a distinct clade with the chrysanthemum cultivars ‘fenhuangxiaori’ and ‘fenzichiguan’ with a 100% bootstrap support value in the *LFY* gene tree (Fig. 2), which suggested that the species was a highly likely ancestor of at least two cultivars. *Chrysanthemum vestitum* did not group with chrysanthemum cultivars into a clade, so we propose that it may have only involved in the origin of chrysanthemums indirectly. Furthermore, the diploid *C. lavandulifolium* formed a clade with three sequences of the tetraploid *C. indicum* with a 100% support bootstrap value (Fig. 2), suggesting that *C. lavandulifolium* may have involved in the origin of *C. indicum* or some of the four cultivars. Our results thus suggest that the different chrysanthemum cultivars originated from various hybridizations between multiple wild species.

In addition, the sequences from the same chrysanthemum cultivar or different cultivars were found in three clades. This pattern demonstrated that hybridization and allopolyploidy may have played an important role in the development of chrysanthemum cultivars. Our results supported that different cultivars of

chrysanthemums may have different donor species, which was consistent with the results of previous work (Li et al. 2011; Liu et al. 2012; Klie et al. 2014).

The sequences from *C. argyrophyllum* were found to be in at least five clades with one clustered with *C. vestitum*, one forming a clade with *C. zawadskii*, and three constituting their own respective clades (Fig. 2). It seems that hexaploid *C. argyrophyllum* may have originated from allopolyploidization involving divergent progenitors. We suggested that *C. vestitum* and *C. zawadskii* may be two of the donors for the genome of *C. argyrophyllum*. The origin of this species seems to be much more complex involving several events of reticulate evolution. This result is also congruent with previous studies (Liu et al. 2012; Ma et al. 2014). More work should be conducted to explore the origin of *C. argyrophyllum* and its reticulate evolutionary relationship with other *Chrysanthemum* taxa.

Based on the evidence from the present *LFY* sequence data in *Chrysanthemum* and previous studies, we provide a summary on the hybridization and polyploidy among various wild donor species in the domestication of the chrysanthemum cultivars (Fig. 3). Our data suggest the involvement of several species with diverse morphology in the hybridization scheme, that has generated the amazing diversity of the cultivated chrysanthemums.

Our results from phylogenetic analysis also showed that the *LFY* gene contains many informative characters and shows low levels of homoplasy. It provided phylogenetic resolution for a number of clades within *Chrysanthemum*. However, it is necessary to incorporate chloroplast sequence data which show maternal inheritance

to explore the origin of the allopolyploid taxa and the donor genomes of chrysanthemums cultivars .

Our analyses included only one sequence from each of the three cultivars, ‘yulupantao’, ‘guandongdaxia’, and ‘xiaoju’. The entire *LFY* sequence is rather long and technically difficult to sequence. We only obtained one positive clone from these three cultivars. The *LFY* sequence of ‘xiaoju’ was clustered into a small clade with *C. nankingense* with a high support value; we thus hypothesized that *C. nankingense* may be one of the ancestors of ‘xiaoju’. The *LFY* sequences of ‘yulupantao’ and ‘guangongdaxia’ were clustered into a clade with *C. indicum*, suggesting that *C. indicum* may be one of their ancestors. Other ancestors of these cultivars clearly need to be explored further in our future efforts.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (No.31470699, 31300158), the Fundamental Research Funds for the Central Universities (No.130420003, QN2012020, 2452015406) and the Laboratories of Analytical Biology of the National Museum of Natural History, the Smithsonian Institution. The China Scholarship Council was gratefully acknowledged for financial support of Yueping Ma’s research visit to the Smithsonian Institution.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Bio* 215: 403-410.
- Blázquez MA, Soowal LN, Lee I, Weigel D (1997) *LFY* expression and flower initiation in *Arabidopsis*. *Development* 124: 3835-3844.
- Blázquez MA, Weigel D (2000) Integration of floral inductive signals in *Arabidopsis*. *Nature* 404: 889-892.
- Chen JY (1985) Studies on the origin of Chinese florist's chrysanthemum. *Acta Hort* 167: 349-361.
- Coen ES, Romero JM, Doyle S, Elliot R, Murphy G, Carpenter R (1990) *Floricaula*: a homeotic gene required for flower development in *Antirrhinum majus*. *Cell* 63: 1311-1322.
- Dai SL, Chen JY (1996) Interspecific hybridization of seven species of *Dendranthema*. *Acta Scent Nat Univ Pekinensis* 18: 16-21(in Chinese with an English abstract).
- Dai SL, Chen JY, Li WB (1998) Application of RAPD analysis in the study on the origin of Chinese cultivated chrysanthemum. *Acta Bot Sin* 40: 1053-1059 (in Chinese with an English abstract).
- Dai SL, Wang WK, Huang JP (2002) Advance of researches on phylogeny of *Dendranthema* and origin of chrysanthemum. *Acta Scent Nat Univ Pekinensis* 24: 230-234 (in Chinese with an English abstract) .
- Dai SL, Wang WK, Xu YX (2005) Phylogenetic relationship of *Dendranthema* (DC.) Des Moul. revealed by fluorescent in situ hybridization. *J Inter Plant Biol* 7:

783-791.

- Darriba D, Taboada GL, Doallo R, Posada D (2012) JModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9: 772. doi:10.1038 / nmeth.2109
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792-1797.
- Frohlich MW, Meyerowitz EM (1997) The search for homeotic gene homologs in basal angiosperms and Gnetales: A potential new source of data on the evolutionary origin of flowers. *Int J Plant* 158: S131-142.
- Frohlich MW, Parker DS (2000) The mostly male theory of flower evolutionary origins: from genes to fossils. *Syst Bot* 25: 155-170.
- Fukai S, Kamigaichi Y, Yamasaki N, Zhang W, Goi M (2002) Distribution, morphological variations and cpDNA PCR-RFLP analysis of *Dendranthema yoshinaganthum*. *J Jap Soc Hort Sci* 71: 114-122.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Mentjies P, Drummond A (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647-1649.
- Klie M, Schie S, Linde M, Debener T (2014) The type of ploidy of chrysanthemum is not black or white: a comparison of a molecular approach to published cytological methods. *Front Plant Sci* 5: 479.40. doi:10.3389/fpls.2014.00479
- Li C, Chen S, Chen F, Li J, Fang W (2011) Cytogenetic study of three edible

Chrysanthemum cultivars. Genetika 47: 199-205.

Liu ZP, Wang NF, Zhao AY, Sheng JH, Liu XL, Liu GS (2007) The evolution and expression of low-copy nuclear genes in allopolyploid plants. Hereditas 29: 163-171(in Chinese with an English abstract).

Liu PL, Wan Q, GuoYP, Yang J, Rao GY (2012) Phylogeny of the genus *Chrysanthemum* L.: Evidence from single-copy nuclear gene and chloroplast DNA sequences. Plos One, e48970

Ma YP, Fang XH, Chen F, Dai SL (2008) *DFL*, a *FLORICAULA/LEAFY* homologue gene from *Dendranthema lavandulifolium* is expressed both in the vegetative and reproductive tissues. Plant Cell Rep 27: 647-654.

Ma YP, Dai SL (2009) Genomic DNA extraction from chrysanthemum using high salt CTAB method. Biotechnol Bulletin 7: 90-93 (in Chinese with an English abstract).

Ma YP, Zhou YZ, Wang YZ, Wei JX, Yu ZY, Yang S, Wang Y, Dai SL (2013a) CnFL, a *FLORICAULA/LEAFY* homolog in *Chrysanthemum nankingense* is dramatically up-regulated in induced shoot apical meristems. Biochem Syst Ecol 50: 114-120.

Ma YP, Zhou YZ, Wei JX, Wang YZ (2013b) Cloning and structure analysis of *LEAFY* homologues of genomic sequence isolated from *chrysanthemum morifolium* 'Qianshouguanyin'. Journal of Northeastern University (Natural), 34: 909-912 (in Chinese with an English abstract).

Ma YP, Wei JX, Yu ZY, Dai SL, Wang YZ (2014) Genetic diversity of *LEAFY*

- homologs in *Chrysanthemum argyrophyllum* L. Biochem Syst Ecol 55: 14-19.
- Peng YY, Wei YM, Baum BR, Yan ZH, Lan XJ, Dai SF, Zheng YL (2010)
Phylogenetic inferences in *Avena* based on analysis of *FL* intron 2 sequences.
Theor Appl Genet 121: 985-1000.
- Ren JB, Guo YP (2015) Behind the diversity: Ontogenies of radiate, disciform, and
discoid capitula of Chrysanthemum and its allies. J Syst Evol doi:
10.1111/jse.12154
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference
under mixed models. Bioinformatics 19: 1572-1574.
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B,
Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian
phylogenetic inference and model choice across a large model space. Systematic
Biol 61: 539-542.
- Rudall PJ, Remizowa MV, Prenner G, Prychid CJ, Tuckett RE, Sokoloff DD (2009)
Nonflowers near the base of extant angiosperms? Spatiotemporal arrangement of
organs in reproductive units of Hydatellaceae and its bearing on the origin of the
flower. Am J Bot 96: 67-82.
- Schultz EA, Haughn GW (1991) *LEAFY*, a homeotic gene that regulates inflorescence
development in *Arabidopsis*. Plant Cell 3: 771-781.
- Swofford DL (2002) PAUP*. Phylogenetic analysis using parsimony (*and other
methods), ver. 4.0b10. Massachusetts: Sinauer Associates, Sunderland.
- Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and

- post-analysis of large phylogenies. *Bioinformatics* 30: 1312-1313.
- Tu T, Dillon MO, Sun H, Wen J (2008) Phylogeny of *Nolana* (Solanaceae) of the Atacama and Peruvian deserts inferred from sequences of four plastid markers and the nuclear *LEAFY* second intron. *Mol Phylogenet Evol* 49: 561-573.
- Wada M, Cao Q, Nobuhiro K, Soejima J, Masuda T (2002) Apple has two orthologs of *FLORICAULA/LEAFY* involved in flowering. *Plant Mol Biol* 49: 567-577.
- Weigel D, Alvarez J, Smyth DR, Yanofsky MF, Meyerowitz EM, 1992. *LFY* controls floral meristem identity in *Arabidopsis*. *Cell* 69: 843–859.
- Winter CM, Austin RS, Blanvillain-Baufume S, Reback MA, Monniaux M, Wu MF, Sang Y, Yamaguchi A, Yamaguchi N, Parker JE (2011) *LEAFY* target genes reveal floral regulatory logic, cis motifs, and a link to biotic stimulus response. *Dev Cell* 20: 430-443.
- Zheng X, Hu C, Spooner D, Liu J, Cao J, Teng Y (2011) Molecular evolution of *Adh* and *LEAFY* and the phylogenetic utility of their introns in *Pyrus* (Rosaceae). *BMC Evol bio* 11: 255. doi:10.1186/1471-2148-11-255
- Zimmer EA, Wen J (2013) Using nuclear gene data for plant phylogenetics: progress and prospects. *Mol Phylogenet Evol* 66: 539-550.

Figure Captions

Fig. 1. Schematic structure of *LFY* gene in species of *Chrysanthemum* (modified from Ma et al. 2014).

Fig. 2. Bayesian 50% majority-rule consensus tree for *LFY* gene in *Chrysanthemum*.

The tree is based on the exon+intron sequences. Posterior probabilities from the Bayesian inference (PP) and bootstrap percentages (BP) from the maximum likelihood analysis are shown next to the branches. Main clades are indicated by I, II and III. Labels of terminal branches include species or variety names followed by clone identities (from 1 to 16).

Fig. 3. Hypotheses on the origin of the chrysanthemum cultivars supported by the *LFY* data. Left was the position of wild species in the *LFY* tree. Right was the different parental donors for the corresponding chrysanthemum cultivars.

Table 1. Plant material used in the experiment and GenBank accessions. Voucher specimens were deposited in the US National Herbarium.

Taxon or cultivar	Voucher specimens	Location in China	No. of clones	GenBank accession no
* <i>C. vestitum</i> (Hemsl.) Stapf	Ma2012HN001	Henan	6	KF142403-KF142408
<i>C. argyrophyllum</i> Ling	Ma2012HN002	Henan	8	KF151337-KF151344
* <i>C. chanetii</i> H. Lev	Ma2012HN003	Henan	6	KF151325-KF151329
* <i>C. zawadskii</i> (Herbich) Tzvelev	Ma2011NJ001	Nanjing	6	KF142397-KF142402
* <i>C. indicum</i> L.	Ma2008HB001	Hebei	8	Submitted
<i>C. nankingemse</i> (Nakai) Tzvel Hand.-Mazz	Ma2011HuB001	Hubei	4	JX123024
<i>C. lavandulifolium</i> (Fisch) (Fisch. ex Trautv.) Makino	Ma2006BJ001	Beijing	3	AY559245
* ¹ <i>C. × morifolium</i> 'xiaoju'	Ma2008BJ001	Beijing	1	JX070180
* <i>C. × morifolium</i> 'guandongdaxia'	Ma2008BJ002	Beijing	1	JX027621
* <i>C. × morifolium</i> 'fenzhouqi'	Ma2008BJ003	Beijing	8	Submitted
* <i>C. × morifolium</i> 'fenhuangxiaori'	Ma2008BJ004	Beijing	8	Submitted
* <i>C. × morifolium</i> 'fenzichiguan'	Ma2008BJ005	Beijing	8	Submitted
* <i>C. × morifolium</i> 'yulupantao'	Ma2008BJ006	Beijing	1	JX027626

Note: * New sequences generated from this study.

¹ Cultivar that has not been named.

Table 2. Information of primers used in *LFY* PCR amplification

Primer name	Primer position	Sequence
LK	Exon 1	ATGGACCCTGATGCACTTTC
LB	3'UTR	ATGGTACCATGGACCCTGATGCACTTTC
GSP 2	Exon 2	GGACCAAGAACTCACGACAT
D1	Exon 2	CCGTTTATCGTGACCGAGCC
GSP 0	Exon 3	CTCATCAAGACAATGCAATGCGT
975	Exon 3	GCTAAGAAGTCTGGTGCAAG
973	Exon 3	CATTCGACCCCTCCTCATCA
110 F	Exon 1	ATGTCGTGAGTTCTTGGTCC
110 R	Exon 2	CTTGCACCAGACTTCTTAGC

Table 3. Sequence characteristics of the *LFY* gene in *Chrysanthemum*

Datasets	Average G+C composition (%)	Aligned length (range)	Number of variable sites/ Number of informative sites
Exon 1	47.3	465 (462-465)	60/33
Intron 1	26.8	1139 (490-885)	155/102
Exon 2	50.4	384	65/42
Intron 2	29.5	1550 (1066-1291)	282/177
Exon 3 and 3'UTR	46	417	79/43
total	36.6	3955 (2887-3348)	641/397

Figure 1

Figure 2

Figure 3