

Gene Duplication and the Evolution of Photoperiodic Flowering in *Helianthus annuus*

Benjamin Blackman¹, Scott Michaels², Loren Rieseberg³

¹Duke University, Box 90338, Durham, NC 27701, USA, bkb7@duke.edu

²Indiana University, 1001 East Third Street, Bloomington, IN 47401, USA, michaels@indiana.edu

³University of British Columbia, 3529-6270 University Blvd, Vancouver, BC V6T 1Z4, Canada, loren.rieseberg@botany.ubc.ca

ABSTRACT

- Although many plants time flowering to photoperiod cues so that they reproduce in optimal seasonal conditions, sunflower was once considered to be solely day neutral. It is now evident, however, that wild and domesticated sunflower varieties actually exhibit tremendous diversity in how flowering time responds to photoperiod. Notably, transitions between photoperiod response types often coincide with important evolutionary events. In our work, we have begun to characterize the genetic changes involved in the evolution of this variation during sunflower domestication and improvement as well as clinal adaptation in wild populations.
- Natural variation in gene expression of sunflower loci homologous to flowering time genes identified in model organisms was assessed in domesticated and wild varieties that differ in photoperiod response. To identify genes specifically involved in the evolution of flowering during domestication, we conducted further genetic mapping, sequence and expression comparison, heterologous complementation, and molecular evolution studies of these sunflower homologs.
- Through this work, we determined that a series of recent duplications produced several copies of the critical flowering time regulator *FLOWERING LOCUS T* in the sunflower genome. A frameshift mutation in one of these copies affects flowering and experienced selection during early domestication. Transgenic studies conducted in *Arabidopsis thaliana* revealed that this mutation likely functions through dominant-negative interference with another paralog. Additional members of this gene family experienced selection during modern breeding when the direction of selection shifted to favor early flowering instead of late flowering. Finally, changes in expression of these paralogs were associated with the evolution of day neutrality and long day response from short day response in domesticated sunflower but only with day neutrality and not long day response in wild sunflower.
- Together, these findings suggest mechanisms through which recent gene duplications may provide the evolutionary raw material for phenotypic innovation and reversibility. This work also demonstrated that independent evolution of similar photoperiod responses in wild and domesticated sunflower has occurred through both similar and distinct genetic mechanisms.
- Our work revealed abundant phenotypic diversity in photoperiodic flowering in wild populations and identified a number of genetic markers associated with flowering time variation in sunflower germplasm. We also identified and functionally characterized a genetic variant that likely contributed to the initial domestication of sunflower by Native American farmers.

Key words: domestication – evolution – flowering time – gene duplication – natural variation – photoperiod

INTRODUCTION

Since day length is often a reliable indicator of seasonal timing, temperate plants frequently use photoperiod cues to trigger or hasten the onset of flowering. Long-day plants flower earlier or only when day length exceeds a minimum threshold; short-day plants flower earlier or only when day length falls below a maximum threshold. Day neutral plants do not exhibit photoperiodic flowering. The genetic pathways through which photoperiod signals are measured and integrated by the plant to promote reproductive development have been well described (Kobayashi and Weigel, 2007). Notably, many of the same genes are involved in both long-day and short-day responses (e.g. Hayama et al., 2003).

Though once reported as strictly day neutral (Allard and Garner, 1940; Almeida and Pereira, 1996; Marc and Palmer, 1981), sunflower is actually the rare species that harbors long-day, short-day, and day neutral accessions. This variation is observed among cultivars (Goyné and Schneiter, 1987) and also among wild populations (Blackman et al., 2011a). To begin to examine the genetic basis for this abundant variation, we characterized a family of sunflower genes homologous to *FLOWERING LOCUS T* (*FT*), a key floral inducer gene regulated by day length (Kardailsky et al., 1999; Kobayashi et al., 1999). Genetic pathways acting in the leaf interpret day length by integrating signals from light and the circadian clock such that active *FT* only accumulates under inductive photoperiods (Valverde et al., 2004). *FT* protein then travels through the phloem to the shoot apical meristem where it interacts with additional genes to initiate a regulatory cascade conferring reproductive meristem identity (Corbesier et al., 2007).

Here, we review and extend our work on the sunflower *FT* homologs. First, we show that the sunflower *FT* paralogs arose by a series of recent lineage-specific duplications. Then, we discuss how these copies were subject to selection at multiple stages of sunflower domestication and how a frameshift mutation in one *FT* homolog, likely a causal adaptive mutation, exerts its phenotypic effect through a new competitive interaction with another *FT* homolog. Finally, we describe how *FT* homolog expression patterns demonstrate that convergent evolution of photoperiod response variation in wild and domesticated sunflower has occurred through both similar and distinct mechanisms.

MATERIAL AND METHODS

Phylogenetic Analysis: Four *FT* homologs were isolated by degenerate PCR and BAC screening (Blackman et al 2010). Protein sequences of these homologs were aligned with homologs from additional angiosperm species with CLUSTALX. Genbank information for all sequences except CesFTL1 (EH783838), CesFTL2 (EH789133), CsFTL1 (AB679270), CsFTL2 (AB679271), and CsFTL3 (AB679272) can be found in Blackman et al., 2010. A maximum-likelihood tree was generated from the alignment using PHYML (Guindon and Gascuel, 2003).

Diversity Panel Sequencing: Partial genomic sequences of the four *FT* paralogs and seven putative neutral loci were obtained from a diversity panel including 18 elite-bred cultivars, 19 landrace cultivars, 23 wild individuals of broad geographic sampling, and 6 *H. argophyllus* plants (Blackman et al., 2011b). Haplotypes were obtained from heterozygotes by sequencing multiple cloned PCR products. Tests for selective sweeps during early domestication or modern improvement were conducted with maximum-likelihood Hudson-Kreitman-Aguade (MLHKA) tests (Wright and Charlesworth, 2004).

Transgenics: Full-length cDNAs of the wild allele of *HaFT1*, the domesticated allele of *HaFT1*, *HaFT2*, and *HaFT3* were amplified and cloned into vector pMDC32 behind the cauliflower mosaic virus 35S promoter by Gateway cloning. Wild-type (Columbia-0) and *ft-1* mutant *A. thaliana* plants were transformed by the floral dip method. Transformation was confirmed by RT-PCR.

Gene Expression Variation: Leaves were collected from plants grown under short days (8 h light/day) and long days (16 h light/day) every four hours over the course of a day. Plants were sampled ~5 weeks post germination. Leaves from three plants were pooled in each of two to three biological replicates per line per time point. RNA was isolated and gene expression was assayed by qRT-PCR on a Stratagene Mx3000P (Blackman et al., 2010). Two to three technical replicates were conducted for each sample.

RESULTS

Phylogenetic analysis of the four *Helianthus annuus FT* (*HaFT*) homolog sequences revealed that these copies arose through a series of recent duplication events, one likely occurring at the base of the Helianthae tribe and the others likely within the genus (Fig. 1). Re-analysis including newly deposited *FT*-like sequences from star thistle and chrysanthemum found that several additional recent lineage specific duplications have occurred within the Asteraceae (Fig. 1).

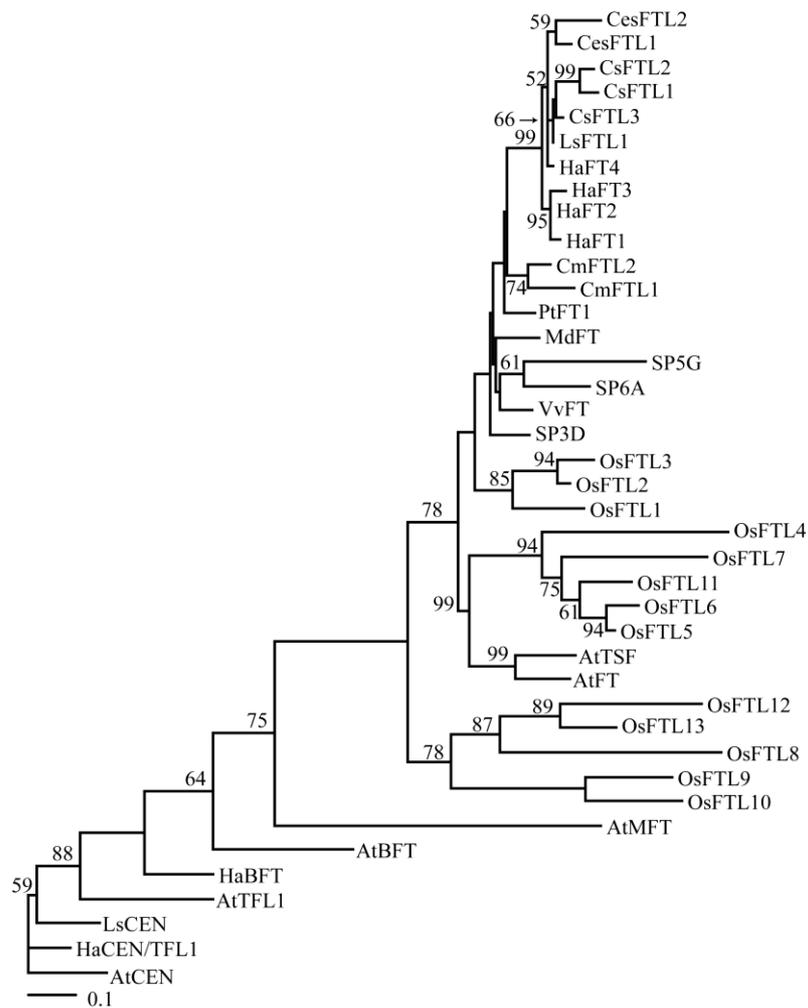


Fig. 1. Multiple lineage-specific duplications of *FT* homologs within the Asteraceae. Bootstrap support is shown for nodes with greater than 50% support. Gene abbreviations are as in Blackman et al., 2010 with the addition of homologs from *Chrysanthemum seticosus* (Cs) and *Centaurea solstitialis* (Ces).

The three recently diverged copies are all closely linked on the distal end of one arm of linkage group 6 (Blackman et al. 2010), co-localizing with a large QTL explaining 7-33% of the variance in flowering time in crosses between wild and cultivated *H. annuus* (Baack et al., 2008; Burke et al., 2002; Wills and Burke, 2007). The final paralog, *HaFT4*, maps to a region on linkage group 14, also near a previous identified flowering time QTL (Baack et al., 2008). These suggestive findings led us to look for coding and expression differences in the *HaFT* paralogs in the parents of one cross, Ann1238 x CMSHA89, and to test for selective sweeps in diversity panel sequences (Blackman et al., 2011b).

We found that *HaFT3* is likely a pseudogene. Its expression was undetectable, and several polymorphisms including a 17.25kb insert, 2 small frameshift-causing indels, and substitutions causing premature stops segregated in the diversity panel (Blackman et al., 2010). *HaFT2* and *HaFT4* were more highly expressed in the cultivated parent than the wild parent (Blackman et al., 2010). *HaFT1* showed two notable changes between the wild and domesticated cross parents (Blackman et al., 2010). First, while only expressed in the wild parent under short days, *HaFT1* was expressed under both short days and long days in the cultivar parent. Second, a frameshift mutation that results in protein longer by 17 amino acids was present in the third exon of the cultivated parent's coding sequence. Compared to neutral loci, all paralogs show reduced levels of sequence diversity in landrace and improved cultigens compared to neutral loci (Fig. 2). MLHKA analysis of diversity panel sequences for the four *HaFT* paralogs provided statistical support for a selective sweeps at *HaFT1* during early domestication and at all four genes during modern improvement (Fig. 2; Blackman et al., 2011b). Evidence for the sweep at *HaFT3* during improvement is likely due to hitchhiking with its closely linked paralogs.

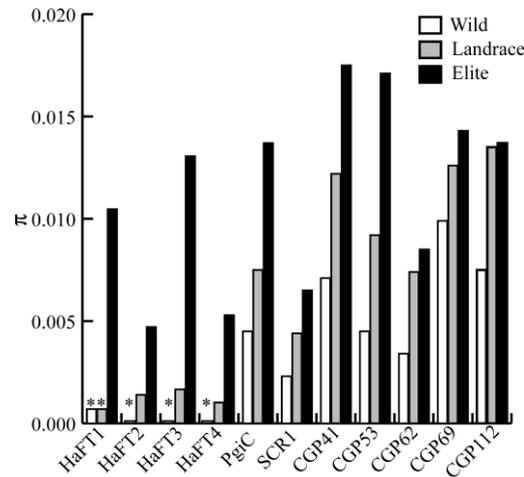


Fig. 2. Reduced pairwise nucleotide diversity (π) in four *HaFT* paralogs with early domestication or improvement. Asterisks note a significant MLHKA test ($p < 0.05$), which is indicative of a signature of a selective sweep. Adapted from Blackman et al., 2010 and Blackman et al., 2011b.

A near isogenic line for the LG6 region segregating for the wild Ann1238 alleles of *HaFT1*, *HaFT2*, and *HaFT3* in the domesticated CMSHA89 background was constructed by five generations of backcrossing. Homozygous domesticated individuals flowered ~7 days later than plant than homozygous wild individuals, but only in long days (Blackman et al., 2010). Since the increased *HaFT2* expression in homozygous domesticated individuals would hasten flowering, we ruled out *HaFT2* as the causal gene. A simple model of a loss of function mutation in a gene with photoperiod-specific expression also did not fit for *HaFT1* since the domesticated allele was expressed under both long and short days. However, we reasoned that *HaFT1* could be the causal mutation if the frameshift mutation produces a dominant negative effect by antagonizing other genes that do only function in long days, namely *HaFT2* or *HaFT4*.

To test this hypothesis, we generated several overexpression lines in the *A. thaliana ft* mutant background: one each for the wild *HaFT1* (*HaFT1-W*), domesticated *HaFT1* (*HaFT1-D*), *HaFT2*, and *HaFT4* cDNAs. All overexpression constructs except *HaFT1-D* rescued the late flowering *ft* phenotype, indicating functional equivalence of these three homologs to *A. thaliana FT* and functional divergence of the domesticated *HaFT1* allele (Fig. 3; Blackman et al., 2010). When double transgenic progeny were generated by crossing the *HaFT1-D* overexpression line to the *HaFT4* overexpression line, rescue of the late flowering phenotype was prevented, indicating a dominant negative interaction (Fig. 3). The same interaction was not observed in plants overexpressing *HaFT1-W* and *HaFT4*. The *HaFT1-W* and *HaFT1-D* constructs differed only by the frameshift mutation; thus the frameshift is causal (Fig 3).

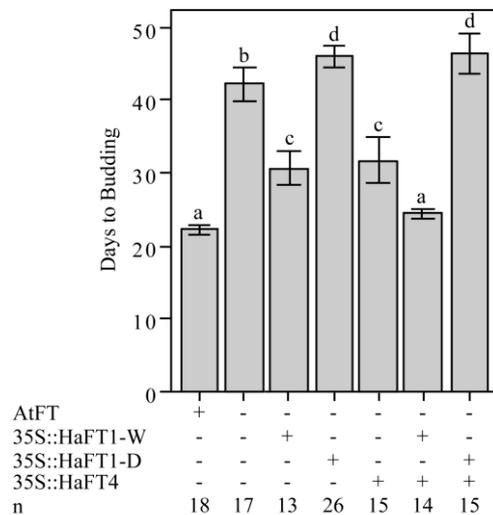


Fig. 3. Dominant-negative interaction between the domesticated allele of *HaFT1* and *HaFT4* demonstrated in *A. thaliana* transgenics. Adapted from Blackman et al., 2010.

To begin to understand how different photoperiod responses have evolved from short-day response multiple times independently in wild and domesticated *H. annuus* (Blackman et al., 2011a; Goynes and Schneiter, 1987), we assayed patterns of *HaFT2* and *HaFT4* expression in day neutral and long day wild individuals and improved cultivars under short and long day conditions (Fig. 4). Wild plants from Manitoba and the cultivated line RHA 274 both exhibit day-neutral, early flowering. In both cases, *HaFT2* was expressed in a typical short-day responsive pattern (Fig. 4A), but *HaFT4* was expressed in both photoperiod conditions (Fig. 4B). Thus, paralog-specific *cis*- or *trans*-acting genetic changes are associated with the evolution of day neutrality in both wild and domesticated sunflower.

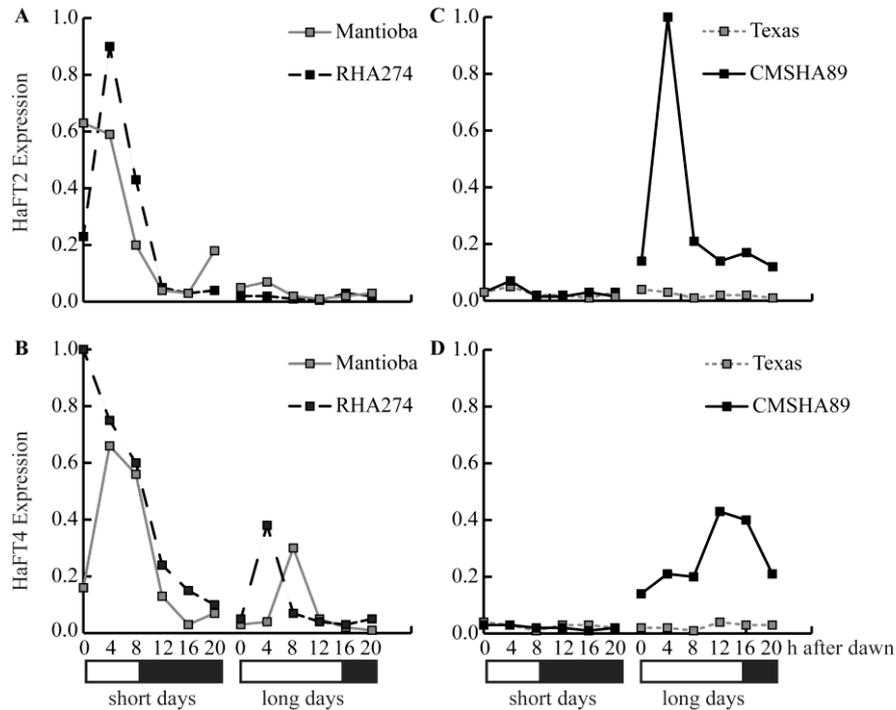


Fig 4. Expression of *HaFT2* and *HaFT4* in day neutral (A and B) and long-day responsive (C and D) wild and domesticated sunflowers. Expression was assayed by qRT-PCR and normalized to *Ha60S rRNA*. Data points represent the mean of 2-3 biological replicates. Since these data are taken from different experiments, expression levels of a given line are shown on a scale relative to the highest expression level measured for that line across both photoperiods.

Wild plants from Texas and the cultivated line CMSHA89 are both long-day responsive, though the Texan plants flower 1-2 months later than the cultivar dependent on photoperiod condition (Blackman et al., 2010, 2011a). While both *HaFT2* and *HaFT4* have evolved to be expressed only under long days in CMSHA89 (Fig. 4C and 4D), neither paralog is expressed above negligible levels under either photoperiod in Texan plants (Fig. 4C and 4D). Additional work has shown that changes affecting *SUPPRESSOR OF CONSTANS 1* homolog expression in the shoot apex are associated with long-day response instead (Blackman et al., 2011a).

DISCUSSION

Our findings demonstrate that recent gene duplications of a critical flowering time regulator have contributed in diverse ways to phenotypic innovations in flowering time and photoperiodic flowering in sunflower. Notably, these changes have been important during the diversification of wild populations and the evolution of cultivated sunflower. The frameshift mutation in *HaFT1* is the first candidate early domestication allele that has been functionally validated for sunflower. Since this functions through a dominant negative interaction with *HaFT4*, these findings also provides some of the first evidence for theoretical models hypothesizing that the evolution of new competitive or cooperative interactions between gene duplicates can elaborate gene networks and lead to adaptive phenotypic change (Pereira-

Leal et al., 2007). Though the *HaFT1* domesticated allele causes later flowering, alleles of *HaFT2* and *HaFT4* putatively causing early flowering were under selection during modern improvement when selection on flowering time likely changed direction. Though those findings require additional functional validation, they suggest that gene duplications can provide a route to evolutionary reversibility. Finally, our comparisons of *HaFT* expression between wild and domesticated accessions demonstrate that both similar and distinct molecular mechanisms can underlie the convergent evolution of similar photoperiodic flowering types.

ACKNOWLEDGEMENTS

We thank J. Strasburg, D. Rasmussen, A. Raduski, K. Turner, J. Burke, S. Knapp, Z. Lai, N. Kane, E. Baack, and the Indiana University Greenhouse staff. Funding in support of this work was provided by the National Science Foundation to B.K.B. (DEB0608118), S.D.M. (IOB0447583), and L.H.R. (DBI0421630) and the National Institutes of Health to S.D.M. (GM075060) and L.H.R. (GM059065).

REFERENCES

- Allard, H.A., and Garner, W.W. 1940. Further observations on the response of various species of plants to length of day. USDA Technical Bulletin 727:35.
- Almeida, J.A.S., and Pereira, M.F.D.A. 1996. The control of flower initiation by gibberellin in *Helianthus annuus* L. (sunflower), a non-photoperiodic plant. *Plant Growth Reg.* 19:109-115.
- Baack, E.J., Sapir, Y., Chapman, M.A., Burke, J.M., and Rieseberg, L.H. 2008. Selection on domestication traits and quantitative trait loci in crop-wild sunflower hybrids. *Mol. Ecol.* 17:666-677.
- Blackman, B.K., Michaels, S.D., and Rieseberg, L.H. 2011a. Connecting the sun to flowering in sunflower adaptation. *Mol. Ecol.* 20:3503-3512.
- Blackman, B.K., Rasmussen, D.A., Strasburg, J.L., Raduski, A.R., Burke, J.M., Knapp, S.J., Michaels, S.D., and Rieseberg, L.H. (2011b). Contributions of flowering time genes to sunflower domestication and improvement. *Genetics* 187, 271-287.
- Blackman, B.K., Strasburg, J.L., Raduski, A.R., Michaels, S.D., and Rieseberg, L.H. 2010. The role of recently derived FT paralogs in sunflower domestication. *Curr. Biol.* 20:629-635.
- Burke, J.M., Tang, S., Knapp, S.J., and Rieseberg, L.H. 2002. Genetic analysis of sunflower domestication. *Genetics* 161:1257-1267.
- Corbesier, L., Vincent, C., Jang, S., Fornara, F., Fan, Q., Searle, I., Giakountis, A., Farrona, S., Gissot, L., Turnbull, C., et al. 2007. FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. *Science* 316:1030-1033.
- Goyne, P.J., and Schneiter, A.A. 1987. Photoperiod influence on development in sunflower genotypes. *Agronomy Journal* 79:704-709.
- Guindon, S., and Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52:696-704.
- Hayama, R., Yokoi, S., Tamaki, S., Yano, M., and Shimamoto, K. 2003. Adaptation of photoperiodic control pathways produces short-day flowering in rice. *Nature* 422:719-722.
- Kardailsky, I., Shukla, V.K., Ahn, J.H., Dagenais, N., Christensen, S.K., Nguyen, J.T., Chory, J., Harrison, M.J., and Weigel, D. 1999. Activation tagging of the floral inducer FT. *Science* 286:1962-1965.
- Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M., and Araki, T. 1999. A Pair of Related Genes with Antagonistic Roles in Mediating Flowering Signals. *Science* 286:1960-1962.
- Kobayashi, Y., and Weigel, D. 2007. Move on up, it's time for change--mobile signals controlling photoperiod-dependent flowering. *Genes & Devel.* 21:2371-2384.
- Marc, J., and Palmer, J.H. 1981. Photoperiodic sensitivity of inflorescence initiation and development in sunflower. *Field Crops Res.* 4:155-164.
- Pereira-Leal, J., Levy, E., Kamp, C., and Teichmann, S. 2007. Evolution of protein complexes by duplication of homomeric interactions. *Genome Biol.* 8:R51.
- Valverde, F., Mouradov, A., Soppe, W., Ravenscroft, D., Samach, A., and Coupland, G. 2004. Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. *Science* 303:1003-1006.
- Wills, D.M., and Burke, J.M. 2007. Quantitative trait locus analysis of the early domestication of sunflower. *Genetics* 176:2589-2599.
- Wright, S.I., and Charlesworth, B. 2004. The HKA test revisited: a maximum-likelihood-ratio test of the standard neutral model. *Genetics* 168:1071-1076.