RESEARCH ARTICLES

Positive Selection and Expression Divergence Following Gene Duplication in the Sunflower *CYCLOIDEA* Gene Family

Mark A. Chapman, James H. Leebens-Mack, and John M. Burke

Department of Plant Biology, Miller Plant Sciences Building, University of Georgia

Members of the *CYCLOIDEA* (*CYC*)/*TEOSINTE-BRANCHED1* (*TB1*) group of transcription factors have been implicated in the evolution of zygomorphic (i.e., bilaterally symmetric) flowers in *Antirrhinum* and *Lotus* and the loss of branching phenotype during the domestication of maize. The composite inflorescences of sunflower (*Helianthus annuus* L. Asteraceae) contain both zygomorphic and actinomorphic (i.e., radially symmetric) florets (rays and disks, respectively), and the cultivated sunflower has evolved an unbranched phenotype in response to domestication from its highly branched wild progenitor; hence, genes related to *CYC/TB1* are of great interest in this study system. We identified 10 members of the *CYC/TB1* gene family in sunflower, which is more than found in any other species investigated to date. Phylogenetic analysis indicates that these genes occur in 3 distinct clades, consistent with previous research in other eudicot species. A combination of dating the duplication events and linkage mapping indicates that only some of the duplications were associated with polyploidization. Cosegregation between *CYC*-like genes and branching-related quantitative trait loci suggest a minor, if any, role for these genes in conferring differences in branching. However, the expression patterns of one gene suggest a possible role in the development of ray versus disk florets. Molecular evolutionary analyses reveal that residues in the conserved domains were the targets of positive selection following gene duplication. Taken together, these results indicate that gene duplication and functional divergence have played a major role in diversification of the sunflower *CYC* gene family.

Introduction

The modification of developmental pathways may provide a powerful substrate for the evolution of morphological diversity (Ohno 1970; Purugganan 1998; Carroll 2000). Transcription factors often play key roles in developmental pathways and hence are likely to be involved in the evolution of morphological variation (Doebley 1993; Doebley and Lukens 1998). In both plants and animals, transcription factors have evolved via gene duplication and functional divergence, giving rise to families of related genes (Scott and Weiner 1984; Purugganan 1998; Moore and Purugganan 2005). Genome-wide investigations suggest that duplicated transcription factors are more commonly retained following polyploidization relative to other classes of genes (Maere et al. 2005). The most well-understood family of transcription factors in plants are the MADS-box genes (reviewed in Yanofsky 1995; Lawton-Rauh et al. 2000; Theissen et al. 2000). However, the TCP family of transcription factors has also received a great deal of attention for its role in regulating differential cell division and floral symmetry in a range of species (Luo et al. 1996, 1999; Cubas, Vincent and Coen 1999; Feng et al. 2006).

Proteins encoded by members of the TCP gene family are characterized by the presence of a basic helix-loop-helix domain called the TCP domain, named after maize *TEOSINTE-BRANCHED1* (*TB1*; Doebley et al. 1997), snapdragon *CYCLOIDEA* (*CYC*; Luo et al. 1996), and rice *PROLIFERATING CELL FACTOR* (*PCF*) *-1* and *-2* (Kosugi and Ohashi 1997). This domain is thought to be involved in DNA-binding and protein–protein interactions (Kosugi and Ohashi 1997, 2002). Twenty-four TCP genes are present in *Arabidopsis*. These genes comprise 2 subfamilies, the *CYC/TB1* subfamily and the *PCF* subfamily, based on amino acid sequence similarity of the TCP domain (Cubas 2002). Within the *CYC/TB1* subfamily, 2 clades of genes are found. One of these, termed the "glutamate-cysteine-glutamate" (ECE) clade by Howarth and Donoghue (2006), contains both *CYC* and *TB1*, and all members of this clade harbor a second conserved domain known as the "R" domain that is thought to mediate protein–protein interactions (Cubas, Lauter et al. 1999). ECE refers to a short region of sequence conservation ECE between the TCP and R domains. Although the second clade contains a subset of genes that also harbor an R domain, it is important to note that this domain was independently recruited within this clade (Cubas 2002). For reasons detailed below, this investigation focuses on the ECE clade of *CYC*- and *TB1*-like genes, which we refer to as "*CYC*-like" genes.

In species that have been investigated to date, there are between 1 (e.g., in grass species; Lukens and Doebley 2001) and 5 (e.g., in the Caprifoliaceae; Howarth and Donoghue 2005) ECE-containing *CYC*-like genes. A large-scale phylogenetic analysis of the *CYC*-like genes from a wide range of flowering plant species revealed the presence of 3 subclades of *CYC*-like genes within the core eudicots, implying that 2 rounds of duplication occurred before diversification of the core eudicots but after the divergence of the core eudicots from the stem eudicots (Howarth and Donoghue 2006).

TB1 is responsible for the difference in branching pattern observed between cultivated maize and its wild progenitor, teosinte (Doebley et al. 1997). In maize, the *tb1* allele is upregulated, increasing apical dominance (i.e., repressing branch elongation) as compared with teosinte. Overexpression of the rice homolog of *TB1* in transgenic rice plants likewise represses branching (Takeda et al. 2003); however, the foxtail millet ortholog of *TB1* has been found to play a relatively minor role in the control of branching in this species (Doust et al. 2004). The *Arabidopsis* gene

Key words: *CYCLOIDEA*, floral development, *Helianthus*, sunflower, *TEOSINTE-BRANCHED1*, transcription factors.

E-mail: mchapman@plantbio.uga.edu.

Mol. Biol. Evol. 25(7):1260–1273. 2008
doi:10.1093/molbev/msn001
Advance Access publication April 3, 2008

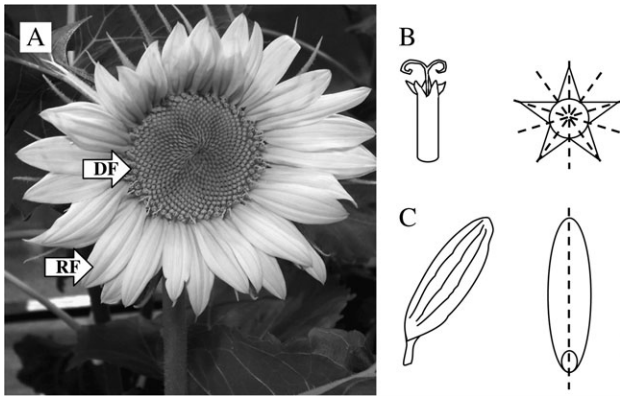

FIG. 1.—(*A*) Flower head of sunflower (*Helianthus annuus*). Disk florets (DF) and ray florets (RF) are labeled. (*B*) DF are actinomorphic, characterized by many planes of symmetry. (C) RF are zygomorphic, characterized by a single plane of symmetry.

BRANCHED1, a possible ortholog of maize *TB1*, plays a similar role as the maize locus in determining plant architecture (Aguilar-Martinez et al. 2007).

In *Antirrhinum majus* (Plantaginaceae) and *Lotus japonica* (Fabaceae), *CYC* is required for the production of zygomorphic (i.e., bilaterally symmetric) flowers (Luo et al. 1996; Feng et al. 2006) via specific expression in the dorsal portion of the flower. Floral symmetry also appears to be controlled in a similar fashion in other species closely related to *Antirrhinum* (Cubas, Vincent and Coen 1999; Hileman et al. 2003). Although not conclusive proof of a causative role, Howarth and Donoghue (2005) found that duplication of 2 *CYC*-like genes occurred at the same time the Caprifoliaceae diverged from other Dipsacales, coinciding with the transition from actinomorphy (i.e., radial symmetry) to zygomorphy. Despite this association, other investigations have failed to establish a direct link between *CYC*-like gene number and/or sequence changes and floral symmetry (e.g., Citerne et al. 2000, 2003; Fukuda et al. 2003; Smith et al. 2004). Zygomorphic flowers have evolved several times in flowering plants (Donoghue et al. 1998), and their origin is thought to have been driven by adaptation to animal pollinator behaviors (Stebbins 1974; Giurfa et al. 1999).

The entire sunflower family (i.e., the Asteraceae or Compositae) is characterized by a composite inflorescence (fig. 1*A*). There is, however, substantial variation in inflorescence architecture among taxa within the family. Indeed, the 3 major subfamilies are distinguished by their inflorescence types: discoid (i.e., composed of actinomorphic disc florets only; the Carduoideae), ligulate (i.e., composed of zygomorphic ray florets only; the majority of the Cichorioideae), and radiate (i.e., composed of both disc and ray florets; the Asteroideae). As a member of the Asteroideae, sunflower (*Helianthus annuus* L.) is therefore characterized by an inflorescence containing both zygomorphic and actinomorphic symmetric florets (fig. 1), and it has been suggested that one or more *CYC*-like genes might be responsible for this developmental difference in radiate species (Gillies et al. 2002). Moreover, cultivated sunflower (also *H. annuus*) has lost the highly branched architecture that is characteristic of its wild progenitor, thereby raising questions about the possible involvement of *CYC*-like genes in this morphological transformation. Given the potential role of *CYC*-like genes in the diversification of flower types in the Asteraceae, we investigated the number, phylogenetic relationships, expression patterns, and genomic locations of members of this gene family in sunflower.

Materials and Methods

Plant Material

Seeds of an inbred line of cultivated sunflower (Ames 3963) were obtained from the United States Department of Agriculture (USDA; http://www.ars-grin.gov/npgs/). This line has previously been used as one of the mapping parents in a quantitative trait loci (QTL) analysis of domestication-related traits in sunflower (see below; Burke et al. 2002). Seeds were clipped to break dormancy, germinated on damp filter paper, transferred to potting compost, and grown to maturity in a greenhouse under 16-h days. Total genomic DNA was isolated from 100 mg of leaf tissue using the DNeasy plant mini kit (Qiagen, Valencia, CA), and total RNA was extracted from root, leaf, ray floret (separated into petals and ovaries), and disc floret (separated into petals, stigmas, and ovaries) of mature plants using the RNeasy kit (Qiagen). Total RNA was treated with RNase-free DNase (Qiagen) to protect against DNA contamination in the extracts.

Polymerase Chain Reaction Amplification and Sequencing

To isolate members of the *CYC* gene family, polymerase chain reaction (PCR) was carried out on DNA using degenerate primers designed from the DNA sequences of the TCP and R domains of *CYC* homologs from various genera (table 1). One primer pair (CTf and CTr; table 1) was designed based on the *CYC* sequence from *A. majus* (Plantaginaceae; GenBank accession number CAA76176) and maize *TB1* (Poaceae; AAB53060). Three further forward primers and 1 reverse primer (CYCf1, CYCf2, CYCf3, and CYCr1; table 1) were designed based on comparisons of *A. majus*, *Linaria vulgaris* (Plantaginaceae; AAD45359), *Populus* × *canescens* (Salicaceae; AAG43046), and *Lupinus nanus* (Fabaceae; AAO88040) and an unpublished sequence from *Senecio squalidus* (Asteraceae; Chapman 2004).

PCR was carried out in 50 μl total volume containing 50 ng DNA, 30 mM tricine pH 8.4–KOH, 50 mM $MgCl_2$, 100 μM of each deoxynucleoside triphosphate, 0.5 μM of each primer, and 2 units of *Taq* DNA polymerase. Primer pairs were used in all 8 possible combinations. Cycling conditions consisted of an initial denaturation for 3 min at 95 °C followed by 35 cycles of 95 °C for 30 s, 50 °C for 30 s, and 72 °C for 1 min and final extension at 72 °C for 20 min. PCR products were resolved on 1.5% agarose gels stained with ethidium bromide. Each combination of primers resulted in at least one band on the gel. PCR products were cloned into pDRIVE PCR cloning vectors (Qiagen), transformed into competent cells, and grown overnight on LB agar plates with antibiotic selection. Ninety-six colonies from each of the 8 initial PCR primer combinations were

Table 1
Primer Sequences Used in This Investigation and the Region of the TCP Domain (Forward Primers) and R Domain (Reverse Primers) for Which the Primers Were Designed

Primer Name	Primer Sequence (5′-3′)	Amino Acid Alignment
CTf	AARGAYAGGCACAGCAA	KDRHSK
CTr	TCCTTRGTYCKYTCCCT	RERTKE
CYCf1	TCTWCAAGABWTGCTAGGKTTYG	LQ(E/D)(M/L)LGF
CYCf2	CAAGABWTGCTAGGKTTYGAYA	Q(E/D)(M/L)LGFD
CYCf3	GARTGGCTYTTTTSCAAGTCYAA	(D/E)WL(F/S/L)(N/D/T)KS
CYCr1	TWGCTCTYGCYCTYGCHT	(A/E)(D/K)ARAR

then screened via PCR using vector primers (T7 and SP6) at 0.2 μM concentration, an annealing temperature of 55 °C, and a total reaction volume of 10 μl. Clones containing an insert of the expected size (250–450 bp) were then treated with 4 units of Exonuclease I and 0.8 units of Shrimp Alkaline Phosphatase (USB, Cleveland, OH) and incubated at 37 °C for 45 min to prepare for sequencing. These products were sequenced in both directions using the T7 and SP6 primers and DYEnamic (Amersham, Piscataway, NJ) sequencing chemistry. Unincorporated dyes were removed by Sephadex (Amersham) cleanup, and the purified products were then sequenced using a BaseStation automated DNA sequencer (MJ Research, San Francisco, CA).

All sequences obtained were analyzed for the presence of an open reading frame and sequence homology with other *CYC*-like genes (limited to the TCP and R domains). All putative *CYC*-like sequences were then compared with each other to identify unique loci. From these, gene-specific primers were designed to amplify the 3′ ends via rapid amplification of cDNA ends (i.e., RACE) using the FirstChoice RLM-RACE kit (Ambion, Austin, TX). The 5′ end of each gene was amplified using either RACE or genome walking (GenomeWalker Universal kit, BD Biosciences, San Jose, CA) following the manufacturers' protocol. RACE was carried out on a 1:1 mixture of ray floret and leaf RNA, whereas genome walking was performed with total genomic DNA.

Once the entire sequence of each gene had been obtained, primers were designed to amplify the full length of each individual copy from genomic DNA. These amplicons were used to confirm the sequences obtained by genomic walking and RACE and verify intron positions via comparison against the cDNA sequences.

Phylogenetic Analyses

To aid in interpreting the relationships between the sunflower paralogs and determine orthology with *CYC*-like genes in other species, a gene tree was constructed. The analysis was restricted to the TCP and R domains because alignments outside these regions were not reliable (see also Reeves and Olmstead 2003), even between some of the more closely related paralogs. The alignment included the 10 genes identified from sunflower plus previously published (Howarth and Donoghue 2006) orthologs and paralogs from the most closely related taxa, including 4 species of the Dipsacales (*Diervilla*, *Lonicera*, *Patrinia*, and *Sambucus*) and 1 species of the Asterales (*Scaevola*). Maximum likelihood (ML) analysis was carried out on the nucleotide alignment using PHYML v2.4.4 (Guindon and Gascuel 2003) under the HKY (Hasegawa et al. 1985) + Γ model of molecular evolution with 4 substitution rate classes. The single *CYC* sequence from *Aquilegia* was used as the outgroup for core eudicot *CYC*–like genes, and bootstrapping was conducted using PHYML with 500 replicates.

Divergence Times and Tests for Selection

To estimate divergence times of the gene family, we carried out ML analysis of the nucleotide sequences of the TCP and R domains of the sunflower *CYC*-like genes. The *Diervilla sessilifolia CYC1* gene (*Ds*CYC1; AY851166) was included in the analysis so as to provide a calibration point. The ML tree was used to estimate divergence times using r8s ver. 1.71 (Sanderson 2003). For this analysis, the node between *DsCYC1* and sunflower *CYC1a* and *CYC1b* sequences was constrained to 94 or 101 MYA following the estimated divergence time of the Asterales and Dipsacales (Wikström et al. 2001). The truncated Newton algorithm of the penalized likelihood method for estimating divergence time was used following recommendations of Sanderson (2003) with an appropriate smoothing parameter (100) estimated by cross-validation (Sanderson 2002).

Patterns of molecular evolution were assessed for the *CYC* gene family using CODEML within the PAML package (v.3.15; Yang 2000; http://abacus.gene.ucl.ac.uk/software/paml.html) and the fitmodel program (v0.5.2; Guindon et al. 2004; http://www.cebl.auckland.ac.nz/~sguindon/fitmodel.html). Our initial alignment was based on the results of BLAST (non-redundant [nr] and database expressed sequence tag [dbEST]) searches (Altschul et al. 1997) and was limited to only *CYC*-like genes that contained the full TCP and R domains. CODEML was used to estimate branch-specific frequencies of synonymous substitutions per synonymous site (dS) across all *CYC* gene family members. We found that the per-site frequency of synonymous substitutions was saturated (dS values >1.0) on many internal branches. Such saturation can confound accurate estimation of selective constraint ($\omega = d\text{N}/d\text{S}$; the ratio of nonsynonymous to synonymous substitutions). Therefore, we limited our analysis of dN/dS to clades within the *CYC1*, *CYC2*, and *CYC3* subfamilies that included all sunflower genes and did not include branches with $d\text{S} > 1$. Separate tests for shifting selective constraint were subsequently run for each subfamily.

A series of likelihood ratio tests (LRTs) were performed to investigate whether some sites were evolving under positive Darwinian selection on all or some of the branches in the *CYC1*, *CYC2*, and *CYC3* gene trees. Tests

were performed for variation in dN/dS across sites and switching from one dN/dS rate ratio class to another across the gene trees (Guindon et al. 2004). Two sets of tests for 3 rate ratio classes were applied: one with a neutral rate ratio class ($\omega = 1$; model M2a of Wong et al. 2004) and the other without any constraint for ML estimation of the 3–rate ratio classes (model M3 with 3–rate ratio classes). Wong et al. (2004; also see PAML user documentation) assert that, whereas the M3 model can be used to test for heterogeneity in selective constraint, the M2a model is more appropriate for identifying signatures of positive selection. Therefore, we first tested for rate heterogeneity by applying the M3 model and then tested for positive selection (M2a) when we found evidence for rate heterogeneity. A 1–rate ratio model (M0) was used as the null hypothesis to test for heterogeneity among sites (M3), and a 2–rate ratio model with $\omega_1 < 1.0$ and $\omega_2 = 1$ (M1a) was used as the null hypothesis for testing positive selection with $\omega_1 < 1.0$, $\omega_2 = 1$, and $\omega_3 > 1$ (M2a).

When we found rate heterogeneity across sites, we tested for switching among rate ratio classes across the gene tree. Guindon et al. (2004) introduced 2 models for switching among rate ratio classes: one with a parameter, δ, specifying the overall switching rates among rate ratio classes (S1) and another with 2 additional parameters, α and β, to allow for unequal switching rates among 3 dN/dS classes (S2). LRTs were used to test for shifting rate ratio across the gene tree (e.g., M2a vs. M2aS1) and unequal switching among rate ratio classes (e.g., M2aS1 vs. M2aS2). When null hypotheses were rejected, the alternative hypothesis was used to estimate posterior probabilities for the assignment of sites to rate ratio classes on each branch, implying significant variation in dN/dS across sites and branches (Guindon et al. 2004).

Expression Analysis

We performed reverse transcriptase–polymerase chain reaction (RT–PCR) using the One-Step RT-PCR kit (Qiagen) to investigate the expression patterns of the 10 *CYC*-like genes in the tissues listed in "Plant Material." Actin was amplified from each tissue as an internal control. Following the reverse transcription reaction (30 min at 50 °C) and polymerase activation (15 min at 95 °C), 35 cycles of PCR were carried out as described above except that the annealing temperature was optimized for each gene (55–63 °C). Products were resolved on agarose gels and sequenced to confirm intron positions.

Genetic Mapping

The sunflower line we used was one of the parents of the wild × cultivated F_3 mapping population used by Burke et al. (2002). We sequenced the 10 genes in a subset of individuals drawn from a recombinant inbred line (RIL) population derived from the original F_3 population. Polymorphisms were detected for 7 of the 10 loci, which allowed these genes to be mapped using PCR–restriction fragment length polymorphisms (RFLPs). Briefly, for each gene in each RIL, PCR amplification was carried out in a 10 μl reaction volume, digested with the appropriate restriction enzyme, and the resultant RFLPs scored from agarose or acrylamide gels. The remaining 3 genes were sequenced and found to be polymorphic in an F_2 population derived from a cross between a primitive sunflower landrace (Hopi) and wild sunflower (Wills and Burke 2007) and were thus mapped in this second population using PCR–RFLPs. Linkage mapping for each population was carried out using MAPMAKER 3.0/EXP (Lander et al. 1987; Lincoln et al. 1992). Recombination fractions were translated into centimorgan distances following Kosambi 1944. Initially, the "group" command in MAPMAKER was used to combine markers with logarithm of the odds >10.0 and $\theta < 0.2$ and marker orders were explored using the "compare" command.

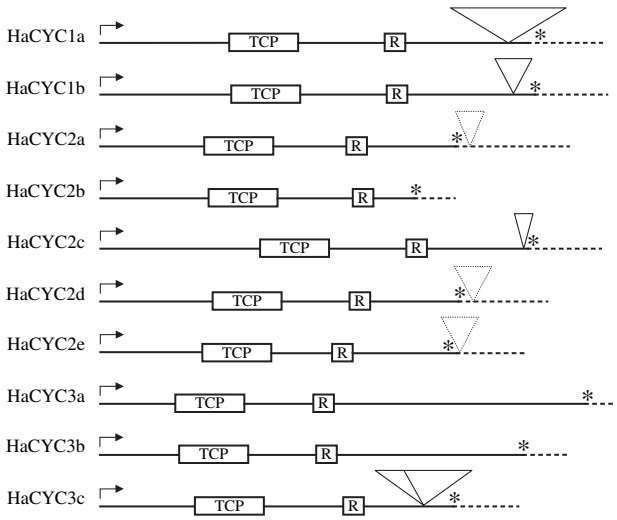

FIG. 2.—Schematic diagram of the predicted mRNAs from 10 *CYC*-like loci in sunflower. All lengths are to scale. The TCP and R domains are of fixed length in all genes and indicated by boxes. The start and stop codons are indicated by an arrow and an asterisk, respectively. Dashed lines indicate the 3′ UTR determined by 3′ RACE. Introns are shown as triangles and are dashed when they are present in the 3 ′ UTR.

Results

The *CYC*-Like Gene Family

Our degenerate PCR approach revealed that sunflower contains 10 members of the *CYC*-like family of TCP transcription factors. The full-length sequences have been deposited in GenBank under accession numbers EU088366–EU088375. Excluding introns, the length of each gene varied from 807 to 1,245 bp (269–415 predicted amino acids), and all 10 contain the expected conserved TCP and R domains (fig. 2). In addition, the 2 *CYC1*-like genes also contained a region very similar to the ECE region of *CYC1* genes in the Dipsacales. Comparison of the genomic and mRNA sequences confirmed that an intron is present in the coding region of 4 genes and in the 3′ untranslated region (UTR) of an additional 3 (fig. 2). *HaCYC3c* appears to be alternatively spliced.

Ancient Duplications in the Sunflower *CYC* Gene Family

The 10 sunflower *CYC*-like genes fall into 3 clades based on ML analysis of the nucleotide sequence of the

TCP and R domains (fig. 3). These 3 clades are consistent with previous studies in other species (Howarth and Donoghue 2005, 2006) and were used as the basis for the naming scheme of the sunflower genes. As observed by Howarth and Donoghue (2006), the *CYC2* and *CYC3* clades are more closely related to each other than to the *CYC1* clade (fig. 3).

Examination of the gene tree shows that gene duplication has played a major role in the evolution of the *CYC*-like gene family in sunflower and, as described previously (Howarth and Donoghue 2006), there has been a complicated pattern of gene duplication and loss within each clade. For example, in the *CYC2* clade, an ancient duplication occurred in the Dipsacales as evidenced by the *CYC2a* and *CYC2b* genes for 3 species of the Dispacales (*Ds*, *Lh*, and *Pt* in fig. 3*A*) and 4 independent duplication events have occurred within the Asteraceae which gave rise to 5 *HaCYC2* genes (fig. 3).

CYC1-Like Genes

In previous studies, only a single copy of *CYC1* has been recovered from most species investigated (Lukens and Doebley 2001; Howarth and Donoghue 2006); however, in the Ranunculales (sister to all other eudicots) as well as in the present study, 2 *CYC1* copies were identified (Kolsch and Gleissberg 2006; Damerval et al. 2007). These duplications appear to have occurred independently as the gene tree shown in figure 3*A* indicates that the *CYC1* duplication in *Helianthus* occurred since the split between the Asterales and Dipsacales. Moreover, previous analyses have indicated that the core eudicot *CYC*–like genes diverged independently of the stem eudicot *CYC*–like genes (Howarth and Donoghue 2006). The expression patterns of these genes have also diverged, with *HaCYC1a* being expressed in all tissues except roots and *HaCYC1b* only expressed in petals (fig. 5).

CYC2-Like Genes

Five of the 10 sunflower *CYC*-like genes are *CYC2*-like genes. Although previous studies have generally found more *CYC2*-like genes than *CYC1*- or *CYC3*-like genes, no more than 3 such genes have ever been reported from other species (Howarth and Donoghue 2006). As described above, phylogenetic analyses suggest independent *CYC2*-like gene duplications within the Asterales and Dipsacales (fig. 3*A*). Further, independent duplication events are evident within the Asterales in lineages leading to *Scaevola* and *Helianthus*.

Within *Helianthus*, there have been several *CYC2*-like gene duplications, with a basal split occurring approximately 44 MYA followed by further duplications (table 2, fig. 4), and these duplications have been followed by expression changes (fig. 5). One gene (*HaCYC2b*) is expressed in every tissue tested and an additional 2 (*HaCYC2a* and *e*) are expressed across all floral parts but not in leaves or roots. The other 2 *CYC2* genes are expressed much more specifically, with *HaCYC2d* expressed in disk petals and ray florets and *HaCYC2c* expressed in ray florets only (in both the petal and ovary). Phylogenetic analyses place *HaCYC2d* and *HaCYC2c* in separate clades, implying independent canalization or expansion of expression domains for sunflower *CYC2* genes. Note that, although *CYC* gene expression has not been reported previously in carpel tissue, the finding that *CYC* expression in the ovary does not appear to be due to either DNA contamination (see above) or contamination of the ovary RNA with that from the petal. Indeed, in the case of petal contamination, we would expect all petal-expressed genes to appear to be ovule expressed, as well. Contrary to this expectation, some genes are expressed in the petal but apparently not in the ovary (e.g., *HaCYC2d*, fig. 5).

CYC3-Like Genes

As seen within the *CYC1*- and *CYC2*-like genes, the 3 *CYC3*-like genes from *Helianthus* were found to be the products of duplication since the divergence of the Asterales from the Dipsacales. Interestingly, the timing of these 3 duplications is very similar at approximately 41–44 MYA. This *CYC3* split is shared with the split between the 2 *Scaevola CYC3* genes (table 2, fig. 4). The 3 *HaCYC3* genes are expressed in most tissues tested (fig. 5); however, *CYC3c* shows the expression of 2 minor transcripts in some tissues (fig. 5) consistent with the occurrence of alternative splicing (see fig. 2).

Map Positions of Duplicate Genes

Some clustering of the *CYC*-like genes in sunflower is evident from the genetic map (fig. 6). Three *CYC2*-like genes are found within 2 cM on LG9, and 2 others (*CYC2a* and *CYC3a*) are found within 3 cM on LG12. The remaining 5 genes are, however, found on separate linkage groups.

Shifting Patterns of Selection in *CYC2* Subfamily

Variation in the mode of selection acting on the TCP and R domains of *CYC1*-, *CYC2*-, and *CYC3*-like genes was assessed via analyses of the ratio of nonsynonymous to synonymous substitution rates (ω). These domains were generally found to be evolving under strong purifying selection, and we could not reject the null hypothesis that all sites in the *CYC1* and *CYC3* alignments were subject to the same high level of constraint ($\omega \leq 0.1$; table 3). In contrast, significant across-site heterogeneity in ω was observed in *CYC2* TCP and R domain alignment. Moreover, we found evidence for shifting modes of selection across the Asteraceae *CYC2* gene tree with adaptive evolution ($\omega > 1.0$) inferred at 4 sites following the first duplication in the *CYC2a*,*b*,*c* clade (table 3, fig. 7). Whereas the LRT failed to detect an improvement in the 3–rate ratio model relative to the 2 class model with one dN/dS ratio fixed at 1.0 (test of model 2a vs. 1a, table 3), models including switching among classes showed a significantly better fit relative to simpler no-switching models (i.e., M3S1 vs. M3 and M2aS1 vs. M2a). Very slight and nonsignificant improvements in likelihoods were observed when switching rates varied depending on rate ratio classes (i.e., M3S2 vs. M3S1 and M2aS2 vs. M2aS1). Positive selection

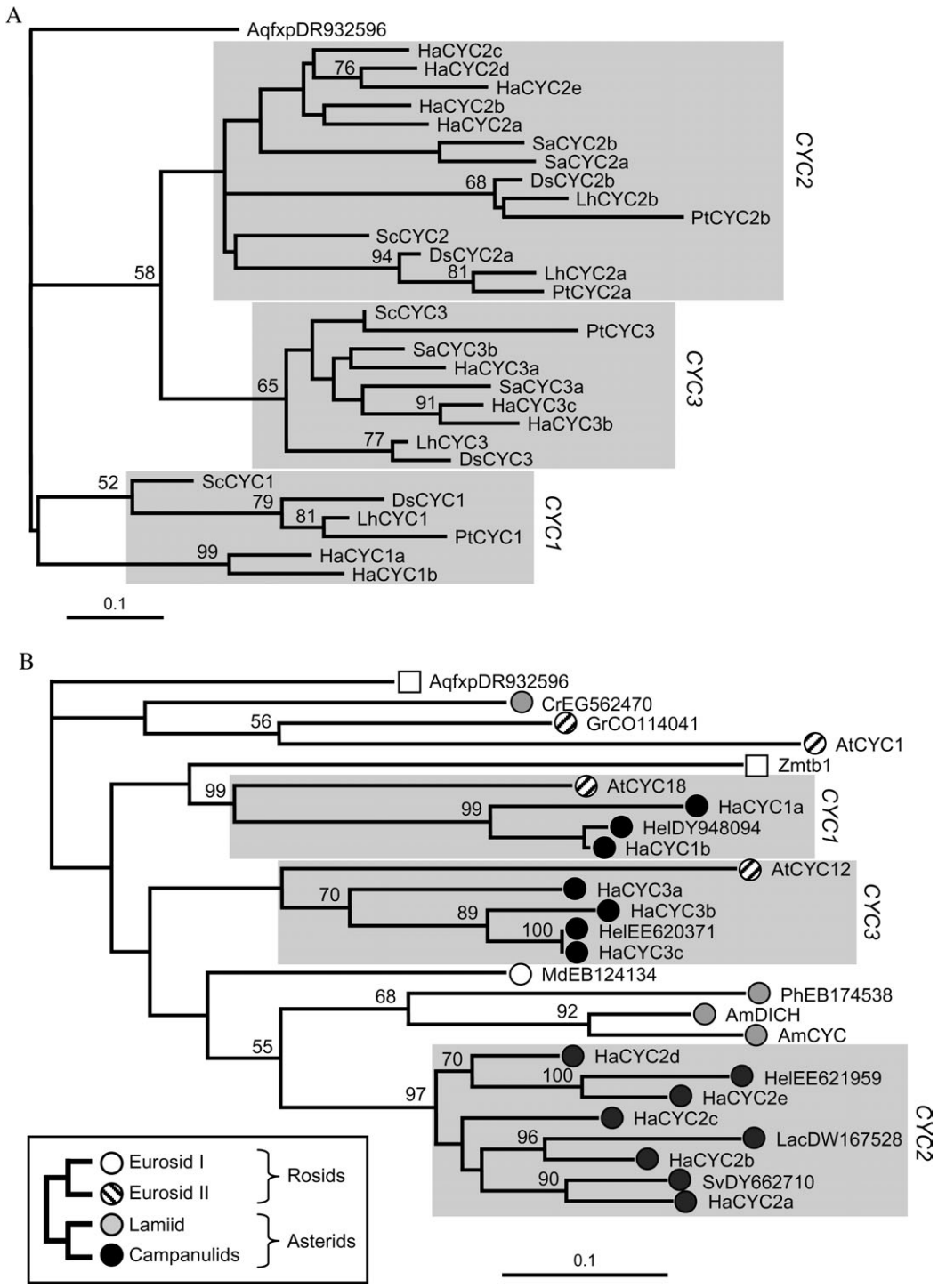

FIG. 3.—ML trees generated by PHYML showing relationships between putative *CYC*-like genes based on the 177-bp TCP domain and 54-bp R domain. The single *CYC*-like gene from *Aquilegia* (Afxp) was used to root both trees. Numbering of each locus follows the original reference or the expressed sequence tag GenBank accession number. Bootstrap percentages are given above branches where >50%. (*A*) To identify orthologs and paralogs of previously published *CYC*-like genes, the 10 novel *CYC*-like genes from *Helianthus annuus* (Ha) were aligned with *CYC*-like genes isolated from *Diervilla sessilifolia* (Ds), *Lonicera heteroloba* (Lh), *Patrinia triloba* (Pt), *Sambucus canadensis* (Sc), and *Scaevola aemula* (Sa). (*B*) For the analysis of selection, the 10 sunflower genes were aligned with sequences from GenBank (nr and dbEST) that comprised the entire TCP and R domains. Species abbreviations are *Antirrhinum majus* (Am), *Aquilegia formosa* × *A. pubescens* (Afxp), *Arabidopsis thaliana* (At), *Catharanthus roseus* (Cr), *Gossypium raimondii* (Gr), *Helianthus* spp. (Hel), *Lactuca* spp. (Lac), *Malus domestica* (Md), and *Zea mays* (Zm). Taxa are labeled phylogenetically according to the inset, and the sequence from *Aquilegia* was used as the outgroup following Howarth and Donoghue (2006). Due to saturation of synonymous substitutions on many long branches, the analyses of variation in selective constraint were limited to the 3 clades indicated by gray boxes.

Table 2
Divergence Time Estimates for the *CYCLODEA*-Like Gene Family in Sunflower

Gene Duplication	Upper Estimate (MYA)	Lower Estimate (MYA)	Mean (MYA)
2/3	70.65	65.75	68.20
3a/3b + 3c	45.18	42.05	43.62
2a + 2b + 2c/2d + 2e	45.07	41.95	43.51
1a/1b	42.55	39.60	41.08
2a + 2b/2c	37.62	35.02	36.32
2d/2e	30.85	28.71	29.78
2a/2b	27.84	25.91	26.88
3b/3c	18.50	17.22	17.86

($\omega > 1.0$) at 3 amino acids in the TCP domain (one each in basic, helix I and helix II) and 1 in the R domain was inferred (posterior probabilities >0.9), regardless of whether 1 of the 3 rate ratio classes was set to neutrality ($\omega = 1.0$; model M2aS1 or S2) or ω was estimated for all 3 of the rate ratio classes (M3S1 or S2). Interestingly, signatures of positive selection for all 4 of these sites are restricted to branches following the *CYC2c*/*CYC2a*,*b* duplication (fig. 7). The ML tree of the *CYC2* genes (figs. 3*B* and 7) showed poor support for the (*CYC2c*, (*CYC2a*, *CYC2b*)) topology, so we repeated the analysis with alternative topologies including clades (*CYC2a*, (*CYC2b*, *CYC2c*)) or (*CYC2b*, (*CYC2a*, *CYC2c*)). In all cases, LRTs supported the positive selection model with varying site-specific heterogeneity in *d*N/*d*S across the tree and equal switching rates among rate ratio classes (data not shown). The same 4 sites showed evidence of positive selection on branches following early diversification within the *CYC2a*/*CYC2b*/*CYC2c* clade. Also noteworthy is the observation that analyses including only the sunflower *CYC2*-like genes (no ESTs) provided significant support for adaptive evolution at the same 4-codon positions (data not shown).

Discussion

The role of *CYC*-like genes in flower development has previously been investigated in a range of species across numerous plant families. Somewhat surprisingly, however, these genes have received little attention in the Asteraceae (but see Gillies et al. 2002; Abbott et al. 2003), which is one of the largest families of flowering plants and one which displays extreme diversity in floral morphology (Funk et al. 2005). A search of the Compositae genome project database reveals that the *CYC*-like genes identified here are not present in the *H. annuus* database. Characterization of *CYC*-like genes in the Asteraceae represents the first step in determining whether or not they play a role in determining floral symmetry in these species, some of which carry zygomorphic and actinomorphic florets within the same inflorescence (e.g., fig. 1).

Gene Structure

The sunflower *CYC*-like genes described herein had similar overall gene structure as compared with previously characterized members of this gene family from other species. Three of the *CYC2*-like genes that we identified were found to have introns in their 3′ UTRs, which has previously been reported for *CYC* in *Antirrhinum* (Luo et al. 1996). In general, UTRs are known to play a central role in gene expression by modulating mRNA localization, stability, and translational efficiency (e.g., Morello et al. 2002; Menossi et al. 2003; Kim et al. 2006; Morello et al. 2006; reviewed in Wilkie et al. 2003 and Hughes 2006), and introns in such regions are relatively uncommon. Moreover, the majority of investigations of UTR-borne introns have focused on those located in 5′ UTRs, presumably because they are more prevalent than introns in 3′ UTRs (Hong et al. 2006). Whereas the function of UTR introns remains relatively poorly understood, the available evidence suggests that they influence expression levels (e.g., Chung et al. 2006; Kim et al. 2006). Interestingly, for one gene (*CYC2e*),

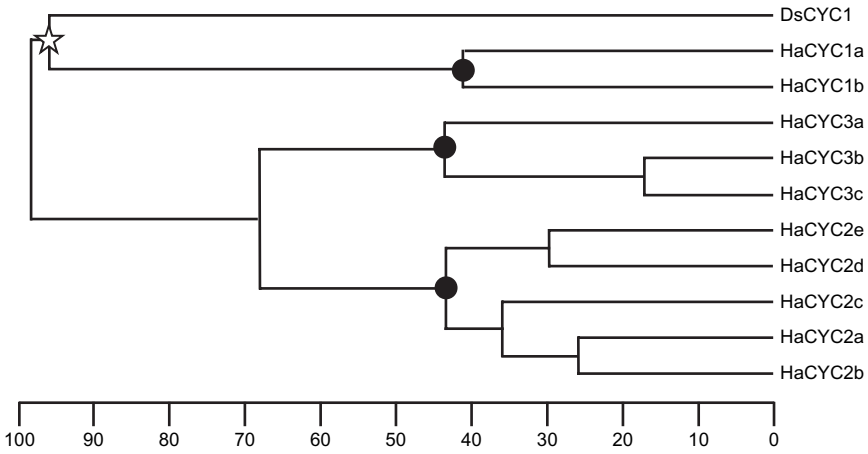

FIG. 4.—Chronogram showing the estimated duplication events within the sunflower *CYC*-like gene family. Scale is in MYA, and mean divergence time is plotted (see table 2). The tree was rooted using the *CYC1* genes. The node indicated with a star was constrained to 94 or 101 MYA. The near-coincident duplications in all 3 lineages are indicated by circles.

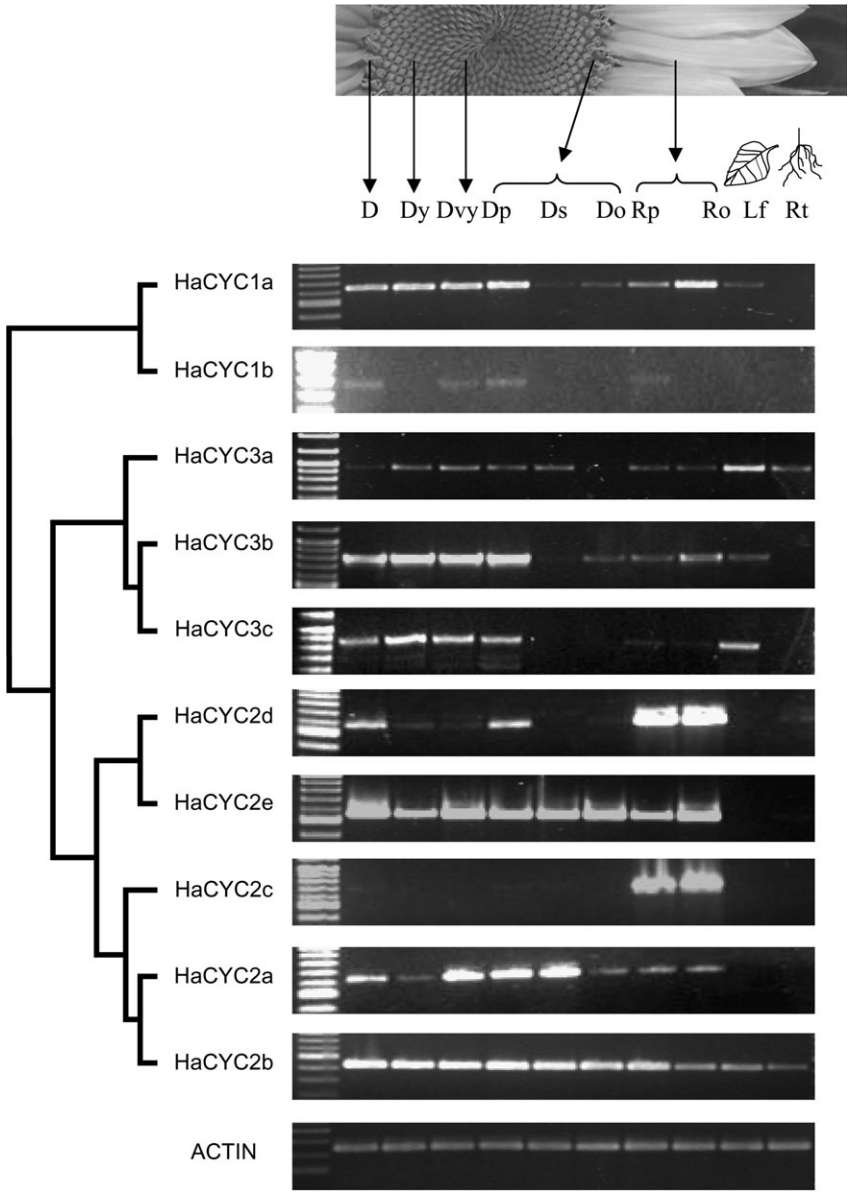

FIG. 5.—Expression of 10 *CYC*-like genes in sunflower following 35 cycles of RT–PCR. The ML tree of the 10 genes is shown on the left. Tissues analyzed: mature disk (D), young disk (Dy), very young disk (Dvy), disk petal (Dp), disk stigma (Ds), disk ovary (Do), ray petal (Rp), ray ovary (Ro), leaf (Lf), and root (Rt).

expression in ray florets appears to result in 2 different transcripts, one of which is unspliced such that it retains the 3′ intron (fig. 5).

Number of Loci and Their Relationships

Previous studies of *CYC*-like genes in a variety of species have revealed the presence of between 1 and 5 genes per species. Monocots and magnoliids contain just 1 *CYC*-like gene (Lukens and Doebley 2001; Howarth and Donoghue 2006), consistently found in the *CYC1* subclade. In the Papaveraceae and Fumariaceae (Ranunculales; stem eudicots), 2 *CYC*-like genes are present, representing a duplication of an ancestral *CYC1*-like gene; as such, these species do not contain *CYC2*- or *CYC3*-like genes (Kolsch and Gleissberg 2006; Damerval et al. 2007). Most species of both the asterids and rosids contain members of all 3 classes of *CYC*-like genes, indicating that the 3 major lineages of *CYC*-like genes arose before the asterid/rosid split, which occurred approximately 120 MYA (Wikström et al. 2001).

Our analysis of *CYC*-like genes in sunflower confirms the presence of the 3 distinct gene lineages in a member of the Asteraceae. Whereas no *CYC1*-like gene was reported in a previous investigation of another member of the Asterales, the authors of that study suggested that this result might be an artifact of difficulties in amplifying *CYC1*-like genes (Howarth and Donoghue 2006). Nonetheless, the discovery of 10 *CYC*-like genes indicates that gene duplication has been a prominent factor in the evolution of this gene

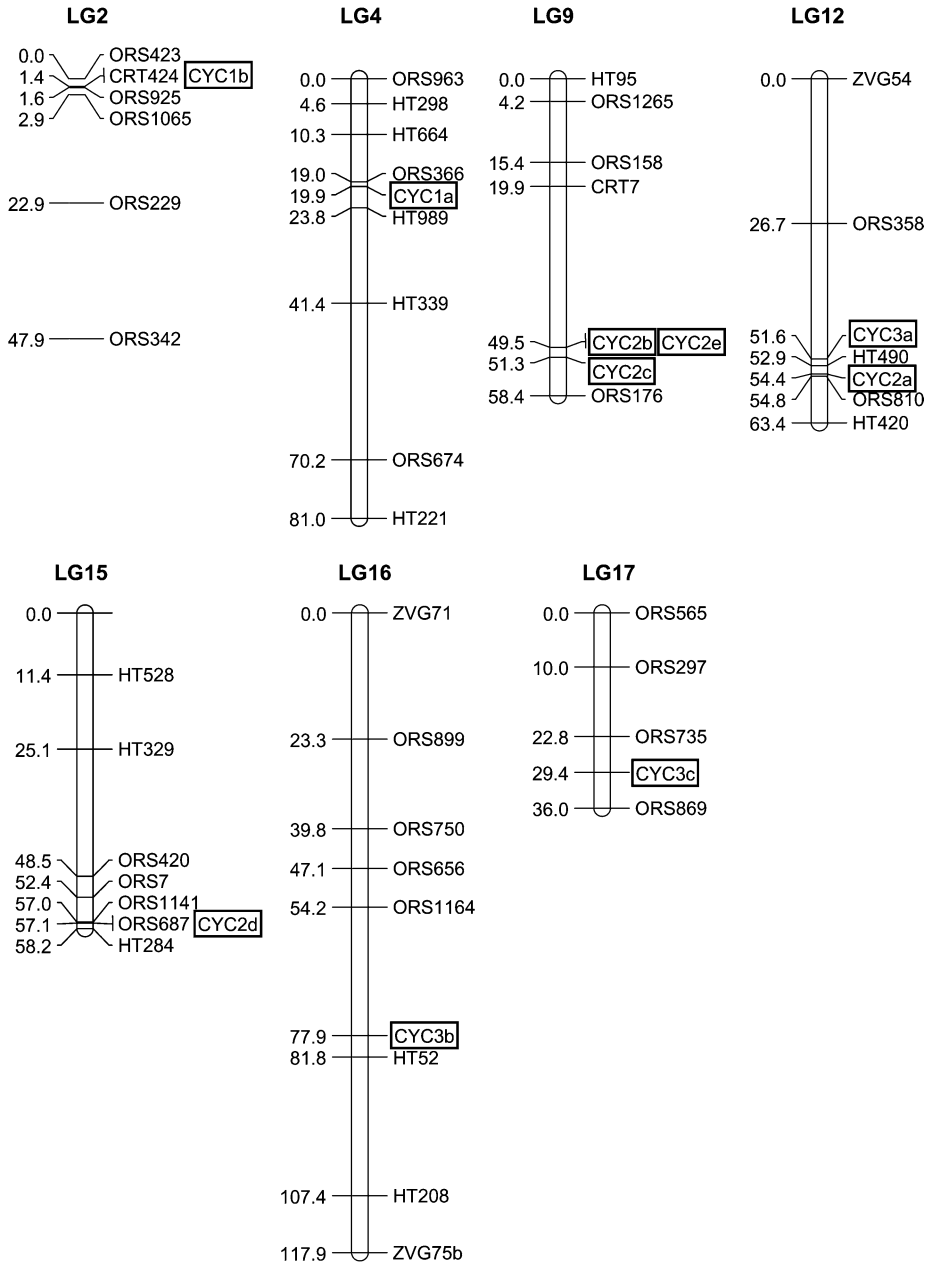

FIG. 6.—Linkage map of sunflower showing map positions of the 10 *CYC*-like genes (boxed). Linkage groups 4, 9, 12, and 16 are taken from a set of 184 RILs, whereas linkage groups 2, 15, and 17 are from 192 hopi × wild F2 plants. See text for details.

family in *Helianthus*, resulting in far more copies than have previously been identified in any taxon studied to date.

Evolution of the *CYC*-Like Gene Family

If we presume that, in all species analyzed thus far, all the *CYC*-like genes that are present have been discovered, then a complex picture of gene loss and duplication begins to emerge. Some general patterns are, however, evident. Most notable among these is the presence of 3 clades of *CYC*-like genes in most core eudicots that have been investigated. Additionally, within each of these 3 clades, all further *CYC* duplications that are evident in *Helianthus* have occurred since the split between the Dipsacales and Asterales (fig. 3).

Although some gene duplication and loss is expected for gene families, it is also likely that the use of degenerate PCR primers to determine the total number of members of a gene family instead provides an estimate of the minimum number of genes within that species (Linhart and Shamir 2005). By incorporating results of a search of the National Center for Biotechnology Information dbEST (http://www.ncbi.nlm.nih.gov/dbEST/), however, we found that all ESTs derived from members of the Asteraceae are closely related to the sunflower *CYC*-like genes (data not shown), providing more (albeit not complete) confidence that other

Table 3
LRTs for Variation in ω

Clade	Model	ln(Likelihood)	LRT Statistic (Models; df; *P* Value)	ω Estimates (Proportion of Sites)
CYC1	M0	−621.044		$\omega_1 = 0.10$ (1.0)
	M3	−619.196	3.696 (M0 vs. M3; 4; 0.44)	$\omega_1 = 0.05$ (0.77); $\omega_2 = 0.35$ (0.12); $\omega_3 = 0.36$ (0.11)
CYC2	M0	−1022.846		$\omega_1 = 0.14$ (1.0)
	M3	−1010.254	25.183 (M0 vs. M3; 4; 4.62E-5)	$\omega_1 = 0.00$ (0.32); $\omega_2 = 0.13$ (0.53); $\omega_3 = 0.65$ (0.15)
	M3S1	−1005.007	10.494 (M3 vs. M3S1; 1; **0.0005**)	$\omega_1 = 0.07$ (0.93); $\omega_2 = 0.10$ (0.04); **$\omega_3 = 3.22$ (0.03)**; $\delta = 0.33$
	M3S2	−1004.283	1.448 (M3S1 vs. M3S2; 2; 0.48)	$\omega_1 = 0.05$ (0.933); $\omega_2 = 1.37$ (0.066); $\omega_3 = 100.00$ (0.001); $\delta = 0.37$; R12 = 5.53; R13 = 65.9; R23 = 1366.32
	M1a	−1010.910		$\omega_1 = 0.09$ (0.88); $\omega_2 = 1.0$ (0.12)
	M2a	−1010.664	0.493 (M1a vs. M2a; 2; 0.77)	$\omega_1 = 0.05$ (0.733); $\omega_2 = 0.46$ (0.261); $\omega_3 = 1.0$ (0.006)
	M2aS1	−1005.023	11.28 (M2a vs. M2aS1; 1; **0.00078**)	$\omega_1 = 0.05$ (0.952); $\omega_2 = 1.0$ (0.002); **$\omega_3 = 3.25$ (0.046)**; $\delta = 0.32$
	M2aS2	−1005.001	0.024 (M2aS1 vs. M2aS2; 2; 0.99)	$\omega_1 = 0.05$ (0.948); $\omega_2 = 1.0$ (0.006); **$\omega_3 = 3.22$ (0.046)**; $\delta = 1.06$; R12 = 73.56; R13 = 1.49; R23 = 115.35
CYC3	M0	−679.458		$\omega_1 = 0.06$ (1.0)
	M3	−678.539	1.837 (M0 vs. M3; 4;0.77)	$\omega_1 = 0.05$ (0.93); $\omega_2 = 0.32$ (0.06); $\omega_3 = 9.58$ (0.01)

NOTE.—Estimated *P* values are based on position of LRT statistic [2 × (log likelihood [H_a] − log likelihood [H_{null}])] in χ distribution with reported degrees of freedom (df) when 2 or greater and a 50:50 mixture of χ0 and χ1 in comparisons with just one extra parameter for the alternate hypothesis (Self and Liang 1987). Values highlighted in bold are indicative of positive selection and are explained in the text.

clades of *CYC*-like genes were not "missed" in our investigation of *Helianthus* due to primer design.

Insights into Sunflower Genome Evolution

Because sunflower is a paleopolyploid (Sossey-Alaoui et al. 1998), one might expect to find evidence of large-scale duplications based on the map positions of closely related members of a gene family such as the *CYC*-like genes. More specifically, barring further rearrangement, pairs of genes resulting from segmental duplication should be found in close proximity to one another, whereas duplication events tracing back to polyploidization should result in pairs of loci split between different chromosomes (e.g., Guillet-Claude et al. 2004; He et al. 2004; Soranzo et al. 2004; but see Kanazin et al. 1996; Michelmore and Meyers 1998; Holland et al. 2000). The map positions of the *HaCYC*-like gene family provide some evidence of large-scale genome duplications, in that the daughter loci following the *CYC1* and *CYC3* duplications approximately 41–44 MYA are found on different linkage groups; however, the situation is less clear for the *CYC2* clade. Whereas 2 lineages (*CYC2d*/*e* vs. *CYC2a*/*b*/*c*) likewise appear to have diverged at the same time, mapping revealed tight linkage associations between some of the more distantly related genes (e.g., *CYC2b* and *e* on LG 9; table 2, fig. 6). It is thus possible that this association is the result of tandem duplications followed by translocations, consistent with the highly dynamic nature of the sunflower genome (Burke et al. 2004). This result, taken alongside the similar divergence estimates for duplication events within the 3 major clades of *CYC*-like genes, supports the occurrence of a polyploidization event approximately 41–44 MYA. This is somewhat earlier than predicted based on the hypothesis that the polyploid event occurred at the base of the Heliantheae sensu lato (Baldwin et al. 2002), which diverged from the remainder of the Asteroideae approximately 20 MYA (Kim et al. 2005).

A Role in Shoot Branching?

As noted above, sunflower domestication involved a loss of the highly branched growth form that is characteristic of wild sunflower, resulting in an unbranched cultivated form topped by a single, large inflorescence. There are, however, some branched breeding lines, and previous work has revealed that this sort of branching results from the effects of either 1 single major locus called *B* (Putt 1964; Tang et al. 2002, 2006) or 2 loci (top- and bottom-branching; Gentzbittel et al. 1999). Despite the apparently simple genetic control of branching in specific cultivar × cultivar crosses, QTL mapping has revealed that branching is under relatively complex control in wild × cultivated sunflower mapping populations (Burke et al. 2002; Wills and Burke 2007).

Although the *B* locus is known to map to LG10 (Tang et al. 2002, 2006), none of the *CYC*-like genes that we identified mapped to this linkage group, suggesting that the *B* locus does not represent a *CYC*-like gene. Similarly, the 2 sunflower *CYC*-like genes that are most similar to maize *TB1* (*HaCYC1a* and *1b*) are found on LG4 and LG2, respectively, and no branching-related QTL have ever been found on either of these linkage groups (Burke et al. 2002; Tang et al. 2002, 2006; Wills and Burke 2007). The *HaCYC3b* and *3c* loci, however, exhibit overlap with branching-related QTL on LG16 and LG17, respectively, in the

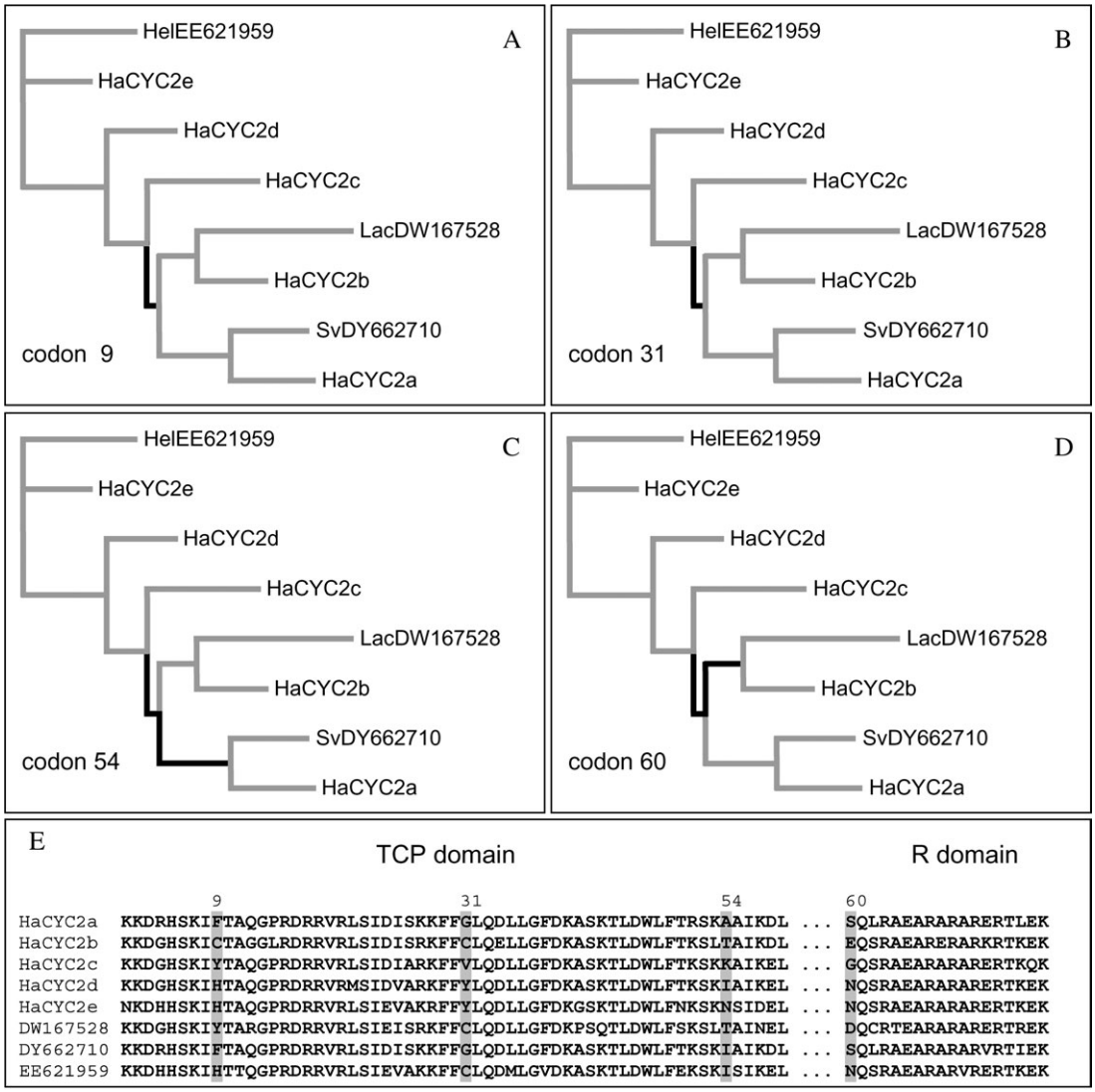

FIG. 7.—Four sites in the TCP and R domains of the *CYC*-like gene family are under positive selection. (*A–D*) The posterior probabilities (pps) for these sites evolving under positive selection ($\omega_3 = 3.25$ in the M2aS1 analysis) are indicated on the trees where black and gray denote pp >0.9 and <0.9, respectively. (*E*) An alignment of the TCP and R domains of *CYC2* sequences from the Asteraceae shows a high degree of variability at amino acid sites 9, 31, 54, and 60 (highlighted).

primitive domesticate × wild F_2 QTL population of Wills and Burke (2007) and also show expression in vegetative tissue (fig. 5). It is therefore possible that one or both of these loci play a minor role in the branching differences that arose during sunflower domestication.

A Possible Role in Floral Symmetry

The finding that one of the *CYC*-like genes (*HaCYC2c*) is only expressed in ray florets is extremely interesting in the context of floral symmetry. As noted above, ray florets are zygomorphic, whereas disc florets are actinomorphic, and it has been suggested that a gene controlling floral symmetry might control the development of ray florets (Gillies et al. 2002). In fact, some species in the Asteraceae are polymorphic for the presence/absence of ray florets and, where it has been investigated, simple genetic control (i.e., 1 or 2 major genes) has been implicated (e.g., *Senecio vulgaris* [Trow 1912]; *Senecio jacobaea* [Andersson 2001]; *Layia* spp. [Ford and Gottlieb 1990]). Until now, *CYC*-like genes have received little attention in the Asteraceae; however, the possible role of a *CYC*-like gene in determining the presence/absence of ray florets is being investigated in *S. vulgaris* (Abbott et al. 2003).

Gene Duplication and Functional Divergence

It has previously been suggested that a common outcome of gene duplication is sub- and/or neofunctionalization (e.g., Lynch and Force 2000; Duarte et al. 2006). Consistent with this view, there is clear evidence of divergence in expression patterns across duplicates within all 3 clades of sunflower *CYC*-like genes. This is most notable in the *CYC2* lineage, in which 3 genes are expressed in all

floral tissues, 1 is restricted to rays and disks, and 1 is restricted to just ray florets. In this clade, we found very strong evidence that positive selection has promoted divergence of the *CYC2a*, *b*, and *c* genes (table 3, fig. 7). Two of the 4 amino acids under selection are found in the helices of the TCP domain and may therefore alter the secondary structure of the proteins. Although the role of the R domain is less well understood than that of the TCP domain, a role in protein–protein interactions has been hypothesized; hence, positive selection on an amino acid in the R domain also hints at functional divergence. When coupled with the notable divergence in expression, it appears that these genes have been under selection for functional divergence. In this context, it is worth noting that Ree et al. (2004) found evidence for positive selection operating on a *CYC* paralog in the genus *Lupinus* that corresponded to a shift in floral morphology, although Hileman and Baum (2003) could not reject the null hypothesis of consistent purifying selection across a comparison of *CYC*-like genes from *Antirrhinum* and its relatives. The possibility of functional divergence in response to divergent selection among paralogs in sunflower could be explored further through experimental investigation of variation at the 4 amino acid sites in the *CYC2* alignment showing signatures of positive selection (e.g., Barkman et al. 2007).

Conclusions

Despite being relatively young (ca. 42–49 MYA; Kim et al. 2005), the Asteraceae is among the largest of plant families with an estimated 24,000–30,000 species (Funk et al. 2005; Stevens 2006), indicating that it has experienced a rapid radiation since its origin. Given the diversity of floral forms within this family, it seems likely that this radiation was driven, at least in part, by floral diversification. Assuming this to be true, the expansion and neofunctionalization of gene families involved in floral development, such as the *CYC*-like genes, may have played a role. The Goodeniaceae are the closest relative of sunflower that has been investigated with respect to *CYC*-like genes and are known to contain 4 *CYC*-like genes (Howarth and Donoghue 2006). Our finding, that sunflower has 10 *CYC*-like genes, indicates that gene duplication has been a prominent factor in the evolution of this gene family. Moreover, following gene duplication events in the *CYC2* lineage, it appears that these genes may have experienced sub- and/or neofunctionalization. It is also noteworthy that, within this clade, only 1 of the 5 genes is expressed outside of floral tissue. *Helianthus* is highly derived within the Asteraceae, and we are currently investigating the evolution of the *CYC*-like genes across the family to determine the timing and phylogenetic position of gene duplications and investigate the relationship between these events and the diverse floral forms that are present in the family.

Acknowledgments

We would like to thank members of the Burke laboratory for comments on an earlier version of the manuscript, Steve Knapp and Bob Brunick for providing the RIL tissue, David Wills for genotypic information from the Hopi × wild mapping population, Dianella Howarth (Yale University) for sequence information, and Thomas Hughes (Leeds University) for information concerning 3′ UTR introns. We also acknowledge the helpful comments of Dianella Howarth and 1 anonymous reviewer. This work was supported by grants from the National Science Foundation (DBI-0332411) and the USDA (03-35300-13104).

Literature Cited

Abbott RJ, James JK, Milne RI, Gillies ACM. 2003. Plant introductions, hybridization and gene flow. Philos Trans R Soc Lond B Biol Sci. 358:1123–1132.

Aguilar-Martinez JA, Poza-Carrión C, Cubas P. 2007. *Arabidopsis BRANCHED1* acts as an integrator of branching signals within axillary buds. Plant Cell. 19:458–472.

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

Andersson S. 2001. The genetic basis of floral variation in *Senecio jacobaea* (Asteraceae). J Hered. 92:409–414.

Baldwin BG, Wessa BL, Panero JL. 2002. Nuclear rDNA evidence for major lineages of helenioid Heliantheae (Compositae). Syst Bot. 27:161–198.

Barkman TJ, Martins TR, Sutton E, Stout JT. 2007. Positive selection for single amino acid change promotes substrate discrimination of a plant volatile-producing enzyme. Mol Biol Evol. 24:1320–1329.

Burke JM, Lai Z, Salmaso M, Nakazato T, Tang S, Heesacker A, Knapp SJ, Rieseberg LH. 2004. Comparative mapping and rapid karyotypic evolution in the genus Helianthus. Genetics. 167:449–457.

Burke JM, Tang S, Knapp SJ, Rieseberg LH. 2002. Genetic analysis of sunflower domestication. Genetics. 161:1257–1267.

Carroll SB. 2000. Endless forms: the evolution of gene regulation and morphological diversity. Cell. 101:577–580.

Chapman MA. 2004. The taxonomy of Senecio sect. Senecio: hybridisation and speciation. St Andrews (UK): University of St Andrews.

Chung BYW, Simons C, Firth AE, Brown CM, Hellens RP. 2006. Effect of 5' UTR introns on gene expression in *Arabidopsis thaliana*. BMC Genomics. 7:120.

Citerne HL, Luo D, Pennington RT, Coen E, Cronk QCB. 2003. A phylogenomic investigation of *CYCLOIDEA*-like TCP genes in the Leguminosae. Plant Physiol. 131:1042–1053.

Citerne HL, Möller M, Cronk QCB. 2000. Diversity of *cycloidea*-like genes in Gesneriaceae in relation to floral symmetry. Ann Bot (Lond). 86:167–176.

Cubas P. 2002. Role of TCP genes in the evolution of key morphological characters in Angiosperms. In: Cronk QCB, Bateman RM, Hawkins JA, editors. Developmental genetics and plant evolution. London: Taylor and Francis Ltd. p. 247–266.

Cubas P, Lauter N, Doebley J, Coen ES. 1999. The TCP domain: a motif found in proteins regulating plant growth and development. Plant J. 18:215–222.

Cubas P, Vincent C, Coen E. 1999. An epigenetic mutation responsible for natural variation in floral symmetry. Nature. 401:157–161.

Damerval C, Le Guilloux M, Jager M, Charon C. 2007. Diversity and evolution of *CYCLOIDEA*-like TCP genes in relation to

flower development in Papaveraceae. Plant Physiol. 143: 759–772.
Doebley J. 1993. Genetics, development and plant evolution. Curr Opin Genet Dev. 3:865–872.
Doebley J, Lukens L. 1998. Transcriptional regulators and the evolution of plant form. Plant Cell. 10:1075–1082.
Doebley J, Stec A, Hubbard L. 1997. The evolution of apical dominance in maize. Nature. 386:485–488.
Donoghue MJ, Ree RH, Baum DA. 1998. Phylogeny and evolution of flower symmetry in the Asteridae. Trends Plant Sci. 3:311–317.
Doust AN, Devos KM, Gadberry MD, Gale MD, Kellogg EA. 2004. Genetic control of branching in foxtail millet. Proc Natl Acad Sci USA. 101:9045–9050.
Duarte JM, Cui LY, Wall PK, Zhang Q, Zhang XH, Leebens-Mack J, Ma H, Altman N, dePamphilis CW. 2006. Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of Arabidopsis. Mol Biol Evol. 23:469–478.
Feng XZ, Zhao Z, Tian ZX, et al. 2006. Control of petal shape and floral zygomorphy in Lotus japonicus. Proc Natl Acad Sci USA. 103:4970–4975.
Ford VS, Gottlieb LD. 1990. Genetic studies of floral evolution in *Layia*. Heredity. 64:29–44.
Fukuda T, Yokoyama J, Maki M. 2003. Molecular evolution of *cycloidea*-like genes in Fabaceae. J Mol Evol. 57:588–597.
Funk VA, Bayer RJ, Keeley S, et al. 2005. Everywhere but Antarctica: using a supertree to understand the diversity and distribution of the Compositae. Biol Skr. 55:343–374.
Gentzbittel L, Mestries E, Mouzeyar S, Mazeryat F, Badaoui S, Vear F, Tourvieille de la Brouhe D, Nicolas P. 1999. A composite map of expressed sequences and phenotypic traits of the sunflower (*Helianthus annuus* L.) genome. Theor Appl Genet. 99:218–234.
Gillies ACM, Cubas P, Coen ES, Abbott RJ. 2002. Making rays in the Asteraceae: genetics and evolution of variation for radiate versus discoid flower heads. In: Cronk QCB, Bateman RM, Hawkins JA, editors. Developmental genetics and plant evolution. London: Taylor & Francis. p. 237–246.
Giurfa M, Dafni A, Neal PR. 1999. Floral symmetry and its role in plant-pollinator systems. Int J Plant Sci. 160:S41–S50.
Guillet-Claude C, Isabel N, Pelgas B, Bousquet J. 2004. The evolutionary implications of *knox-I* gene duplications in conifers: correlated evidence from phylogeny, gene mapping, and analysis of functional divergence. Mol Biol Evol. 21:2232–2245.
Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52:696–704.
Guindon S, Rodrigo AG, Dyer KA, Huelsenbeck JP. 2004. Modeling the site-specific variation of selection patterns along lineages. Proc Natl Acad Sci USA. 101:12957–12962.
Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol. 22:160–174.
He LM, Du CG, Covaleda L, Xu ZY, Robinson AF, Yu JZ, Kohel RJ, Zhang HB. 2004. Cloning, characterization, and evolution of the NBS-LRR-encoding resistance gene analogue family in polyploid cotton (*Gossypium hirsutum* L.). Mol Plant Microbe. 17:1234–1241.
Hileman LC, Baum DA. 2003. Why do paralogs persist? Molecular evolution of *CYCLOIDEA* and related floral symmetry genes in Antirrhineae (Veronicaceae). Mol Biol Evol. 20:591–600.
Hileman LC, Kramer EM, Baum DA. 2003. Differential regulation of symmetry genes and the evolution of floral morphologies. Proc Natl Acad Sci USA. 100:12814–12819.
Holland N, Holland D, Helentjaris T, Dhugga KS, Xoconostle-Cazares B, Delmer DP. 2000. A comparative analysis of the plant cellulose synthase (CesA) gene family. Plant Physiol. 123:1313–1323.
Hong X, Scofield DG, Lynch M. 2006. Intron size, abundance, and distribution within untranslated regions of genes. Mol Biol Evol. 23:2392–2404.
Howarth DG, Donoghue MJ. 2005. Duplications in *CYC*-like genes from dipsacales correlate with floral form. Int J Plant Sci. 166:357–370.
Howarth DG, Donoghue MJ. 2006. Phylogenetic analysis of the "ECE" (*CYC/TB1*) clade reveals duplications predating the core eudicots. Proc Natl Acad Sci USA. 103:9101–9106.
Hughes TA. 2006. Regulation of gene expression by alternative untranslated regions. Trends Genet. 22:119–122.
Kanazin V, Marek LF, Shoemaker RC. 1996. Resistance gene analogs are conserved and clustered in soybean. Proc Natl Acad Sci USA. 93:11746–11750.
Kim KJ, Choi KS, Jansen RK. 2005. Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae). Mol Biol Evol. 22:1783–1792.
Kim MJ, Kim H, Shin JS, Chung CH, Ohlrogge JB, Suh MC. 2006. Seed-specific expression of sesame microsomal oleic acid desaturase is controlled by combinatorial properties between negative cis-regulatory elements in the SeFAD2 promoter and enhancers in the 5′-UTR intron. Mol Genet Genomics. 276:351–368.
Kolsch A, Gleissberg S. 2006. Diversification of *CYCLOIDEA*-like TCP genes in the basal eudicot families Fumariaceae and Papaveraceae s.str. Plant Biol. 8:680–687.
Kosambi DD. 1944. The estimation of map distances from recombination values. Ann Eugen. 12:172–175.
Kosugi S, Ohashi Y. 1997. PCF1 and PCF2 specifically bind to cis elements in the rice proliferating cell nuclear antigen gene. Plant Cell. 9:1607–1619.
Kosugi S, Ohashi Y. 2002. DNA binding and dimerization specificity and potential targets for the TCP protein family. Plant J. 30:337–348.
Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L. 1987. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics. 1:174–181.
Lawton-Rauh AL, Alvarez-Buylla ER, Purugganan MD. 2000. Molecular evolution of flower development. Trends Ecol Evol. 15:144–149.
Lincoln S, Daly M, Lander E. 1992. Constructing genetic maps with MAPMAKER/EXP 3.0. 3rd ed. Cambridge (MA): Whitehead Institute.
Linhart C, Shamir R. 2005. The degenerate primer design problem: theory and applications. J Comput Biol. 12:431–456.
Lukens L, Doebley J. 2001. Molecular evolution of the teosinte branched gene among maize and related grasses. Mol Biol Evol. 18:627–638.
Luo D, Carpenter R, Copsey L, Vincent C, Clark J, Coen ES. 1999. Control of organ asymmetry in flowers of *Antirrhinum*. Cell. 99:367–376.
Luo D, Carpenter R, Vincent C, Copsey L, Coen E. 1996. Origin of floral asymmetry in *Antirrhinum*. Nature. 383:794–799.
Lynch M, Force AG. 2000. The origin of interspecific genomic incompatibility via gene duplication. Am Nat. 156:590–605.
Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y. 2005. Modeling gene and genome duplications in eukaryotes. Proc Natl Acad Sci USA. 102:5454–5459.
Menossi M, Rabaneda F, Puigdomenech P, Martinez-Izquierdo JA. 2003. Analysis of regulatory elements of the

promoter and the 3' untranslated region of the maize Hrgp gene coding for a cell wall protein. Plant Cell Rep. 21:916–923.
Michelmore RW, Meyers BC. 1998. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res. 8:1113–1130.
Moore RC, Purugganan MD. 2005. The evolutionary dynamics of plant duplicate genes. Curr Opin Plant Biol. 8:122–128.
Morello L, Bardini M, Cricri M, Sala F, Breviario D. 2006. Functional analysis of DNA sequences controlling the expression of the rice OsCDPK2 gene. Planta. 223:479–491.
Morello L, Bardini M, Sala F, Breviario D. 2002. A long leader intron of the Ostub16 rice beta-tubulin gene is required for high-level gene expression and can autonomously promote transcription both in vivo and in vitro. Plant J. 29:33–44.
Ohno S. 1970. Evolution by gene duplication. New York: Springer.
Purugganan MD. 1998. The molecular evolution of development. Bioessays. 20:700–711.
Putt ED. 1964. Recessive branching in sunflowers. Crop Sci. 4:444–445.
Ree RH, Citerne HL, Lavin M, Cronk QCB. 2004. Heterogeneous selection on *LEGCYC* paralogs in relation to flower morphology and the phylogeny of *Lupinus* (Leguminosae). Mol Biol Evol. 21:321–331.
Reeves PA, Olmstead RG. 2003. Evolution of the TCP gene family in Asteridae: cladistic and network approaches to understanding regulatory gene family diversification and its impact on morphological evolution. Mol Biol Evol. 20:1997–2009.
Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol Biol Evol. 19:101–109.
Sanderson MJ. 2003. R8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics. 19:301–302.
Scott MP, Weiner AJ. 1984. Structural relationships among genes that control development—sequence homology between the Antennapedia, Ultrabithorax, and Fushi Tarazu loci of *Drosophila*. Proc Natl Acad Sci USA. 81:4115–4119.
Self SG, Liang KY. 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions. J Am Stat Assoc. 82:605–610.
Smith JF, Hileman LC, Powell MP, Baum DA. 2004. Evolution of GCYC, a Gesneriaceae homolog of CYCLOIDEA, within Gesnerioideae (Gesneriaceae). Mol Phylogenet Evol. 31:765–779.
Soranzo N, Gorla MS, Mizzi L, De Toma G, Frova C. 2004. Organisation and structural evolution of the rice glutathione S-transferase gene family. Mol Genet Genomics. 271:511–521.
Sossey-Alaoui K, Serieys H, Tersac M, Lambert P, Schilling E, Griveau Y, Kaan F, Berville A. 1998. Evidence for several genomes in Helianthus. Theor Appl Genet. 97:422–430.
Stebbins GL. 1974. Flowering plants: evolution above the species level. Cambridge: Harvard University Press.
Stevens PF. 2006. Angiosperm Phylogeny Website, Version 7 [Internet]. [cited 2006 May] http://www.mobot.org/MOBOT/research/APweb
Takeda T, Suwa Y, Suzuki M, Kitano H, Ueguchi-Tanaka M, Ashikari M, Matsuoka M, Ueguchi C. 2003. The *OsTB1* gene negatively regulates lateral branching in rice. Plant J. 33:513–520.
Tang S, Yu JK, Slabaugh MB, Shintani DK, Knapp SJ. 2002. Simple sequence repeat map of the sunflower genome. Theor Appl Genet. 105:1124–1136.
Tang SX, Leon A, Bridges WC, Knapp SJ. 2006. Quantitative trait loci for genetically correlated seed traits are tightly linked to branching and pericarp pigment loci in sunflower. Crop Sci. 46:721–734.
Theissen G, Becker A, Di Rosa A, Kanno A, Kim JT, Munster T, Winter KU, Saedler H. 2000. A short history of MADS-box genes in plants. Plant Mol Biol. 42:115–149.
Trow AH. 1912. On the inheritance of certain characters in the common groundsel—*Senecio vulgaris*—and its segregates. J Genet. 2:239–276.
Wikström N, Savolainen V, Chase MW. 2001. Evolution of the angiosperms: calibrating the family tree. Proc R Soc Lond B Biol Sci. 268:2211–2220.
Wilkie GS, Dickson KS, Gray NK. 2003. Regulation of mRNA translation by 5′- and 3′-UTR-binding factors. Trends Biochem Sci. 28:182–188.
Wills DM, Burke JM. 2007. QTL analysis of the early domestication of sunflower. Genetics. doi: 10.1534/genetics.107.075333.
Wong WSW, Yang ZH, Goldman N, Nielsen R. 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics. 168:1041–1051.
Yang Z. 2000. Phylogenetic analysis by maximum likelihood (PAML). London: University College.
Yanofsky MF. 1995. Floral meristems to floral organs—genes controlling early events in Arabidopsis flower development. Annu Rev Plant Physiol. 46:167–188.

Douglas Crawford, Associate Editor

Accepted December 28, 2007