

# Evolution of the TCP Gene Family in Asteridae: Cladistic and Network Approaches to Understanding Regulatory Gene Family Diversification and Its Impact on Morphological Evolution

Patrick A. Reeves<sup>1</sup> and Richard G. Olmstead

Department of Botany, University of Washington, Seattle

In the plant subclass Asteridae, bilaterally symmetrical flowers have evolved from a radially symmetrical ancestral phenotype on at least three independent occasions: in the Boraginaceae, Solanaceae, and Lamiales. Development of bilateral flower symmetry has been shown to be determined by the early-acting *cycloidea* (*cyc*) and *dichotoma* (*dich*) genes in *Antirrhinum*, a member of the Lamiales. *cyc* and *dich* belong to the TCP gene family of putative transcription factors. TCP gene sequences were isolated from 11 Asteridae taxa using an array of degenerate PCR primers. Closely related species exhibiting either ancestral actinomorphic or derived zygomorphic flowers were sampled for each independent origin of bilateral flower symmetry. Cladistic and network-based analyses were performed to establish viable hypotheses regarding the evolution of bilateral symmetry in Asteridae. For the TCP gene family, the use of cladistic phylogenetic analysis to identify orthologous genes is complicated by a paucity of alignable data, frequent gene duplication and extinction, and the possibility of reticulate evolution via intergenic recombination. These complicating factors can be generalized to many regulatory gene families. As an alternative to cladistic analysis, we propose the use of network analysis for the reconstruction of regulatory gene family phylogenetic and functional relationships. Results of analyses support the hypothesis that the origin of bilaterally symmetrical flowers in the Boraginaceae and Solanaceae did not require orthologs or functional analogs of *cyc* or *dich*. This suggests that the genetic mechanism that determines bilateral flower symmetry in these taxa is not homologous to that of the Lamiales. Results of analyses are consistent with the hypothesis that the evolution of bilateral floral symmetry in the Lamiales required the origin of a novel gene function subsequent to gene duplication.

## Introduction

One of the most fundamental problems in modern evolutionary biology is the origin of morphological novelty. Advances in molecular genetics have permitted this problem to be redefined as a question of whether changes in gene content, gene expression pattern, or gene function have been responsible for the evolution of morphological differences between organisms. Within the flowering plants, interest in the genetic mechanisms underlying morphological evolution has been spurred by an increased understanding of early flower development in the model organisms *Antirrhinum majus* and *Arabidopsis thaliana* (Schwarz-Sommer et al. 1990; Coen and Meyerowitz 1991). Numerous single-gene mutations have been identified in these taxa that result in mutant phenotypes suggestive of the evolutionarily-derived morphological differences seen between angiosperm taxa. These include mutations that affect floral organ number, flower symmetry, and floral sex expression (reviewed in Coen 1991).

Flower symmetry has been studied extensively by plant developmental biologists. Three genes have been identified in the model organism *Antirrhinum majus*, a member of the order Lamiales, which act differentially along a dorsal/ventral axis, resulting in the development of bilaterally symmetrical flowers (Luo et al. 1996; Almeida, Rocheta, and Galego 1997). Of these, the action of the

related *cycloidea* (*cyc*) and *dichotoma* (*dich*) genes are best understood. *Antirrhinum cyc/dich* double mutants do not exhibit retarded rates of development in dorsal regions of the floral meristem like the wild-type, and, as a result, flowers are radially symmetrical at maturity. Expression of the *cyc* gene in the wild-type occurs very early in development in dorsal regions of the flower, consistent with its hypothesized action as a repressor of growth (Luo et al. 1996). *dich* has also been shown to influence the early establishment of bilateral symmetry, but to a lesser degree. The primary action of the *dich* gene appears to occur later in development, ensuring the elaboration of dorsal/ventral asymmetry within petals (Luo et al. 1999).

*cyc* and *dich* belong to the TCP gene family, named for genes characterized in *Zea mays* (*tb1*), *Antirrhinum majus* (*cyc*), and *Oryza sativa* (*pcf1*) (Cubas et al. 1999). Members of the gene family contain a highly conserved region, the TCP domain, which forms a noncanonical basic helix-loop-helix structure and functions in DNA binding and protein-protein dimerization (Kosugi and Ohashi 1997, 2002). This function is consistent with the hypothesized role of TCP genes as transcriptional regulators of cell division and growth (Cubas et al. 1999). Phylogenetic analysis of major TCP gene lineages suggests that *tb1*, *cyc*, and *tcp1* (from *Arabidopsis thaliana*) form a subfamily (referred to here as the *cyc/tb1* subfamily) that can be distinguished from other class II function TCP genes by a unique conserved region, the R domain, which is predicted to form a hydrophilic  $\alpha$ -helix (Cubas et al. 1999; Kosugi and Ohashi 2002).

## Evolution of Bilateral Symmetry in the Asteridae

Within the plant family Asteridae, bilaterally symmetrical (monosymmetric, zygomorphic) flowers have arisen independently from radially symmetrical (polysymmetric,

<sup>1</sup> Present address: USDA-ARS, National Center for Genetic Resources Preservation, Fort Collins, Colorado.

Key words: development, evolution, gene family, flower symmetry, *cycloidea*, phylogenetics, network.

E-mail: reevesp@lamar.colostate.edu.

*Mol. Biol. Evol.* 20(12):1997–2009, 2003

DOI: 10.1093/molbev/msg211

*Molecular Biology and Evolution*, Vol. 20, No. 12,

© Society for Molecular Biology and Evolution 2003; all rights reserved.

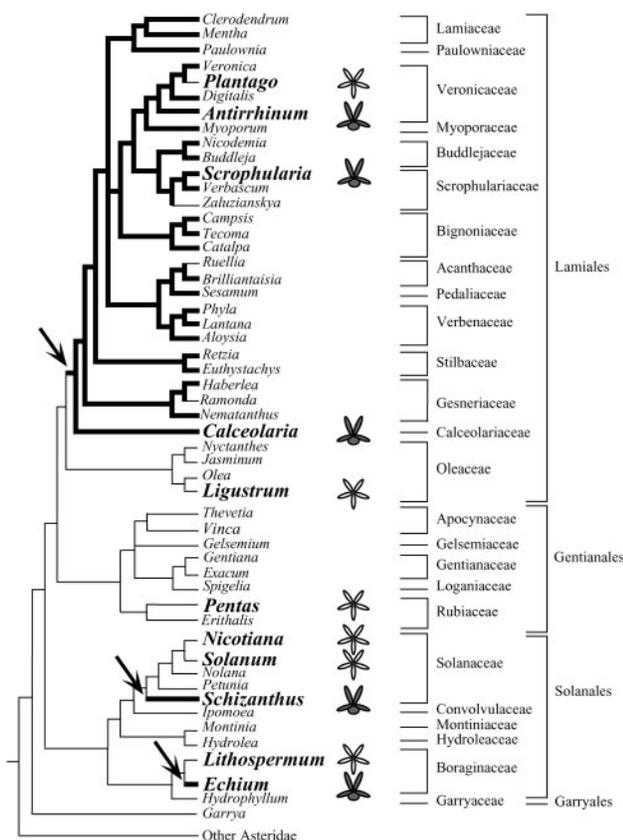


FIG. 1.—Phylogenetic relationships among select Asteridae taxa (modified from Olmstead et al. 2000). Bold branches indicate taxa with bilaterally symmetrical flowers. Bilateral or radial floral symmetry of sampled taxa (bold text) is emphasized by symbols to their right. Arrows indicate three historically independent origins of bilateral symmetry inferred in the Asteridae by Coen and Nugent (1994).

actinomorphic) ancestors in the Boraginaceae, Solanales, and Lamiales (Olmstead et al. 1993; Coen and Nugent 1994). Within the Lamiales, the pattern of dorsal retardation during early floral development is similar among many species (Endress 1999). Given that the expression pattern of *cyc* in *Antirrhinum* is localized in dorsal regions and that a *cyc* homolog has been shown to be involved in the establishment of bilateral symmetry in another member of the Lamiales, *Linaria* (Cubas, Vincent, and Coen 1999), it is likely that a *cyc* homolog and a *cyc*-like expression pattern are required for the development of bilaterally symmetrical flowers throughout the Lamiales. However, it is not clear whether *cyc* homologs are required for development of zygomorphic flowers in the Boraginaceae or Solanales or whether loss of *cyc* function is a prerequisite for the reversion to radial symmetry that has occurred on several occasions in the Lamiales (Baum 1998; Reeves and Olmstead 1998; Endress 1999; Citerne, Möller, and Cronk 2000).

The multiple independent origins of bilateral floral symmetry in the Asteridae may be explained by the following: (1) a change in gene content (i.e., the evolution of new genes via gene duplication), (2) modification of an ancestral regulatory gene expression pattern. This may occur through alteration in upstream regulators or change

in *cis*-acting regulatory sequences of the candidate gene. (3) Modification of an ancestral regulatory gene function. This may occur through change in DNA binding specificity.

To build an experimental and analytical framework to distinguish between these three possible genetic mechanisms underlying morphological evolution, this study examines the pattern of gene evolution within the *cyc/tb1* subfamily of TCP genes in three independent origins of bilateral flower symmetry within subclass Asteridae, where taxonomic relationships are well understood (Olmstead et al. 2000; Albach et al. 2001) (fig. 1).

#### Analytical Methods for Understanding Morphological Evolution

For genetic studies of morphological change, if the candidate gene (*cyc* in the present case) belongs to a gene family, it is necessary to be able to identify the members of the gene family most appropriate for comparisons among taxa. Two methods may be used. First, phylogenetic analysis of DNA sequences can be used to identify putative orthologs. Characterization of orthologous gene expression pattern and function can reveal the genetic changes that must have occurred during the evolution of a morphological difference. Second, estimation of putative functional diversity within a gene family may be used to identify groups of genes with hypothetically similar functions among taxa or to highlight differences in the content of inferred protein functions between taxa. This method does not require inference of historical relationships between the genes sampled.

#### The Cladistic-Historical Approach

The use of cladistic analysis (and other tree-building methods) to identify orthologous genes is complicated by several factors. Many of the transcription factors that regulate key steps in early flower development, such as the MADS box and TCP genes, belong to diverse gene families (Doyle 1994; Purugganan et al. 1995; Theissen, Kim, and Saedler 1996; Cubas et al. 1999; Vieira, Vieira, and Charlesworth 1999). Errors stemming from the use of gene families for reconstructing species phylogenies have been well described (Maddison 1997; Page and Charleston 1997; Martin and Burg 2002). These concerns apply equally well to the identification of orthologs within gene families using cladistic methodologies. Gene duplication, extinction, and intergenic recombination (via processes such as gene conversion) greatly increase the probability of misidentification of orthologs in a cladogram (Sander and Doyle 1992). In addition, genes that are deemed to be members of a family based on conservation of the typically short DNA-binding domain often have highly divergent sequences elsewhere in the transcribed sequence (Atchley and Fitch 1997; Cubas et al. 1999; Rosinski and Atchley 1999; Moore et al. 2000). Consequently, the amount of alignable DNA sequence necessary to accurately reconstruct a gene genealogy may not be attainable for many regulatory gene families.

The reflection of an accepted organismal phylogeny among members of a gene family within a gene tree is the

**Table 1**  
**DNA Source Information and GenBank Accession Numbers**

Taxon name	DNA Source/Voucher Location	GenBank accession numbers
<i>Antirrhinum majus</i>	USDA PI 613013; P. A. Reeves 26 (WTU)	AY168140-AY168144
<i>Calceolaria tenella</i>	S. J. Wagstaff 92-001 (WTU)	AY168152-AY168155
<i>Echium plantagineum</i>	USDA PI 288260; P. A. Reeves 25 (WTU)	AY168175-AY168177
<i>Ligustrum ovalifolium</i>	P. A. Reeves 13 (WTU)	AY168156-AY168158
<i>Lithospermum multiflorum</i>	R. G. Olmstead 93-69 (WTU)	AY168178-AY168184
<i>Nicotiana sylvestris</i>	Verne Sisson, Oxford Tobacco Research Station, TW 138	AY168163-AY168164
<i>Pentas lanceolata</i>	R. G. Olmstead s.n. (WTU)	AY168159-AY168162
<i>Plantago major</i>	University of Washington Medicinal Herb Garden, in cultivation	AY168138-AY168139
<i>Schizanthus aff. pinnatus</i>	Ed Hume Seed; P. A. Reeves 24 (WTU)	AY168168-AY168174
<i>Scrophularia californica</i>	R. G. Olmstead 93-89 (WTU)	AY168145-AY168151
<i>Solanum lycopersicum</i>	Territorial Seed TM784; P. A. Reeves 23 (WTU)	AY168165-AY168167

principal evidence used to assert orthology. Simulation studies have shown that accurate reconstruction of phylogenetic relationships requires sizable amounts of sequence data (>500 variable sites) when branch lengths are variable (Huelsenbeck and Hillis 1993), as would be expected when considering ancient gene families. An inability to accurately reconstruct phylogenetic relationships among sampled taxa within a gene genealogy due to an inherently limited number of characters is a severe impediment to ortholog identification.

#### *The Network-Historical Approach*

Phylogeneticists and population biologists have noted a number of ways in which the assumptions of cladistic methodologies are violated when attempting to reconstruct historical relationships at the intraspecific level from gene sequence information (reviewed by Posada and Crandall 2001). Within sexually reproducing species, relationships among genes have both a hierarchical (or phylogenetic) component and a nonhierarchical (or tokogenetic) component resulting from recombination. The nonhierarchical component cannot be adequately represented by the bifurcating Steiner tree (or “cladogram”) resulting from traditional cladistic analyses. Therefore, whenever genetic recombination is a possibility, the use of spanning trees, or “networks,” to depict historical relationships has been recommended (Templeton, Crandall, and Sing 1992; Excoffier and Smouse 1994; Fitch 1997; Bandelt, Macaulay, and Richards 2000; Smouse 2000; Legendre and Makarenkov 2002; Posada and Crandall 2002). This includes multigene families, where progenitor-derivative and reticulate relationships are possible.

#### *The Functional Approach*

If reticulation has occurred during the evolution of some members of a gene family, it is not logical to attempt to identify orthologs using cladistic analysis of DNA sequences. A chimeric gene contains sequence from two or more ancestral genes. Therefore, it cannot be claimed that it, as a unit, is an ortholog of any one gene in related taxa.

For the same reason, orthologs also may not be identified when network methods are used to reconstruct historical relationships.

Further experimental investigation into the genetic mechanisms underlying morphological change can also be guided using a functional, rather than historical, criterion for characterizing gene family diversity. In the present case, given that the TCP domain determines DNA binding and dimerization specificity (Kosugi and Ohashi 2002), analysis of amino acid sequences could, in principle, be used to identify functional analogs (defined here as sequence variants that, based on measures of amino acid genetic distance, are likely to have identical DNA-binding and dimerization specificities, such that their *trans*-regulatory potential is identical). These sequences would be expected to be capable of exerting very similar regulatory influences, regardless of species of origin, provided that they are expressed in an identical pattern and appropriate downstream target genes are present. Furthermore, functional analogs might be expected to be capable of complementing one another in transgenic experiments. The identification of functional analogs of a candidate gene in two morphologically distinct species may suggest that a difference in expression pattern is responsible for the difference in morphology. Alternatively, the observation that morphologically distinct species do not contain functional analogs within the candidate gene family suggests that a difference in gene content may be responsible.

Here we show that a functional approach may be preferable to historical approaches by documenting the inability of cladistic analysis to provide robust hypotheses of orthology within the TCP gene family of Asteridae. We follow this by demonstrating the use of network analysis of functional diversity to develop hypotheses regarding the evolution of bilateral symmetry in Asteridae.

#### **Materials and Methods**

Closely related species exhibiting either the ancestral radially symmetrical state or the derived bilaterally



**Table 3**  
**Success Rates for Degenerate Primer Pairs Used to Amplify *cyc/tb1* Subfamily TCP Gene Sequences**

Primer Pair	TCP Sequences Recovered	TCP Recovery Rate <sup>a</sup>	Sequence Variants Found	Taxa Where Primer Pair Was Successful
<i>cycA2/9320.1c</i>	1	0.3%	1	1
<i>cycA2/noAt.1c</i>	1	0.3%	1	1
<i>cycA2/Rdom.1c</i>	111	30.0%	27	10
<i>cycA3/9320.1c</i>	4	1.1%	3	2
<i>cycA3/noAt.1c</i>	19	5.1%	4	2
<i>cycA3/Rdom.1c</i>	7	1.9%	1	1
<i>tcA2/9320.1c</i>	78	21.1%	10	6
<i>tcA2/noAt.1c</i>	63	17.1%	6	4
<i>tcA2/Rdom.1c</i>	85	23.0%	15	7

<sup>a</sup> Percentage of the total number of TCP sequences found that are attributable to a particular primer pair.

Because of this, the term “sequence variant” has been used to describe sequences that are believed to be biologically distinct, without reference to their status in the genome. Sequence variants were discriminated from one another and from variants produced by Taq error as follows. Putative TCP sequences were aligned for each taxon individually using ClustalX (Jeanmougin et al. 1998). Aligned sequences were manually divided into groups based on identity in sequence and length. A sequence variant was defined whenever a group of cloned sequences could be distinguished from all other groups in the alignment based on at least one of the following two criteria: (1) presence of a unique indel that did not alter reading frame; (2) presence of two or more unique nucleotides relative to the most similar related group of sequences from the clone pool.

The latter criterion was used infrequently because the high rate of sequence evolution in the variable region between the TCP and R domains permitted identification of sequence variants based on indels. However, the number of sequence variants identified for certain taxa may be elevated slightly due to the use of this second criterion. Sequence variants defined using this criterion were found to cluster during analyses, so use of the second criterion does not impact any conclusions.

Taq errors were eliminated from each group of sequence variants by the creation of majority-rule consensus sequences. This may have caused the number of variants to be underestimated in some taxa but did not affect conclusions. Recombinant sequences generated during PCR (Bradley and Hillis 1997; Cronn et al. 2002) were easily identified as low-frequency (0.5% experiment wide) chimeras of two defined sequence variants and were discarded.

#### Data Set Assembly

Amino acid translations of all variant sequences were aligned with the characterized genes *tb1*, *tcp1*, *cyc*, and *dich* using ClustalX (Jeanmougin et al. 1998). *tcp2* and *tcp3* were included as outgroups as suggested by Cubas et al. (1999). Due to the high rate of sequence evolution between the TCP and R domains, only the TCP domain could be aligned confidently across all taxa and paralogs. The DNA sequence alignment was based on the amino acid sequence alignment.

#### Cladistic Analysis

The probability of long branches due to high sequence divergence prompted the use of maximum-likelihood (ML) analysis for gene genealogy reconstruction (Felsenstein 1978). Likelihood parameters were estimated using the Akaike information criterion as implemented in Modeltest version 3.06 (Posada and Crandall 1998) and used as settings in PAUP\* version 4.0b10 (Swofford 1999). One hundred replicate ML searches were performed using random taxon addition, the heuristic search option, and TBR branch swapping. Bootstrap support values for the ML tree were estimated from 131 replicates using ML as the optimality criterion and the same DNA substitution model as for the ML tree search. ML analyses were performed on a cluster of 20 Macintosh G3 computers.

#### Gene Duplication Rate

The number of gene duplications necessary to account for the diversity of sequence variants observed in the sampled taxa was estimated for all ML trees using GeneTree version 1.0 (Page and Charleston 1997). The species phylogeny used for evaluating gene duplication rate was synthesized from published molecular systematic studies (Chase et al. 1993; Olmstead and Reeves 1995; Olmstead et al. 2000). Polytomies in the ML trees were resolved manually in a manner that minimized the number of inferred duplications.

#### Network Analysis

A distance matrix was calculated for the amino acid alignment of the TCP DNA-binding domain using the Jones, Taylor, and Thornton (1992) amino acid substitution model as implemented in the ProtDist module of PHYLIP version 3.6a2 (Felsenstein 2001). Because the aligned region required no gaps, the amino acid substitution model used by ClustalX to generate the alignment (Gonnet 250) was irrelevant to the calculation of the distance matrix. This distance matrix was used to generate a network of putative functional diversity using three different approaches: the minimum spanning network (Minspnet [Excoffier 1993]), the reticulogram (T-REX [Makarenkov 2001]), and the splitsgraph (SplitsTree [Huson 1998]). Distance-based network reconstruction methods have been used because genetic distances

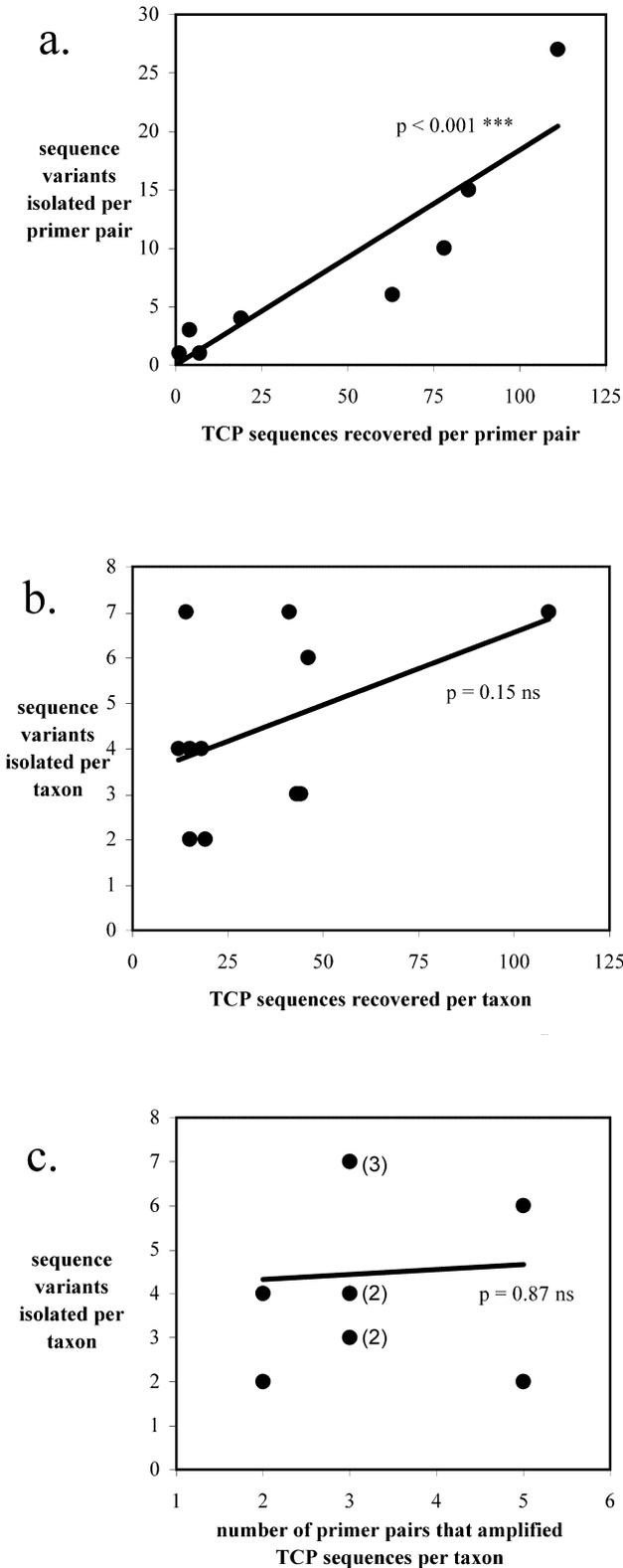


FIG. 3.—Regression analyses pertaining to bias in the amplification of TCP genes using a battery of degenerate primers. (a) Primer pairs that amplified a greater number of TCP gene sequences also produced a greater number of sequence variants. Larger scale studies could be optimized to minimize the number of primer pairs necessary by focusing on those that generate the greatest proportion of the total sequence variants in a pilot study. (b) There is no significant linear relationship between the total

calculated using an appropriate model of amino acid substitution are likely to be more accurate indicators of functional divergence than the stepwise results recovered from unweighted, character-based methods. The splits-graph was largely unresolved and is not discussed further.

**Results**

Clone Screening

Putative TCP sequences were found in 376 clones (GenBank accession numbers AY166131 to AY166169, AY166185 to AY166346, and AY166365 to AY166423). Fifty TCP sequence variants were identified, 47 of which belonged to the *cyc/tbl* clade. Recovery rates are shown in table 2.

The *cyc* gene (here named *AmTCP2*) as well as two probable *dich* sequence variants (*AmTCP3* and *AmTCP4*) from *Antirrhinum* were found. Although it is likely that the two sequence variants are alleles at the *dich* locus (they differ by seven nucleotides), we cannot exclude the possibility that they represent distinct loci. In conflict with the Southern blot data of Luo et al. (1996), which suggests that there are two *cyc*-like genes in *Antirrhinum* (*cyc* and *dich*), Vieira, Vieira, and Charlesworth (1999) have gathered evidence that there may be as many as five *cyc*-like loci (including *dich*) in *Antirrhinum*. The results of our study only confirm the presence of two loci, *cyc* and *dich*, but cannot exclude the possibility of one additional *cyc*-like locus.

Primer Performance

Of the nine pairwise primer combinations used, four primer pairs (tcA2/9320.1C, tcA2/noAt.1C, tcA2/Rdom.1C, and *cyc*A2/Rdom.1C) were responsible for amplifying 91% of the TCP sequences found. The same four primer pairs produced 85% of the sequence variants belonging to the *cyc/tbl* subfamily that were found experiment-wide (table 3). A significant linear relationship between the number of clones attributable to a particular primer pair and the number of TCP sequence variants recovered was found (fig. 3a).

To address the potential for biases in TCP sequence variant recovery, a series of analyses exploring primer pair utilization were undertaken. The results of regression analyses suggest that (1) TCP sequence diversity present in the clone pools was adequately sampled (fig. 3b), and (2) TCP sequence diversity recovered was representative of

number of TCP sequences recovered from a clone pool and the number of sequence variants found. This suggests that screening a greater number of clones for TCP sequences would not increase the number of variants found; thus, clone pools were adequately sampled. (c) In taxa where more primer pairs were effective at amplifying TCP gene targets, a greater number of sequence variants were not predictably found. In the absence of gene lineage-specific amplification bias, this suggests that adding more primer pairs is not likely to increase the number of sequence variants recovered. Given that such bias was found (for *cyc*A2/Rdom.1C), it is not clear whether adding additional primer pairs would result in the recovery of more sequence variants.

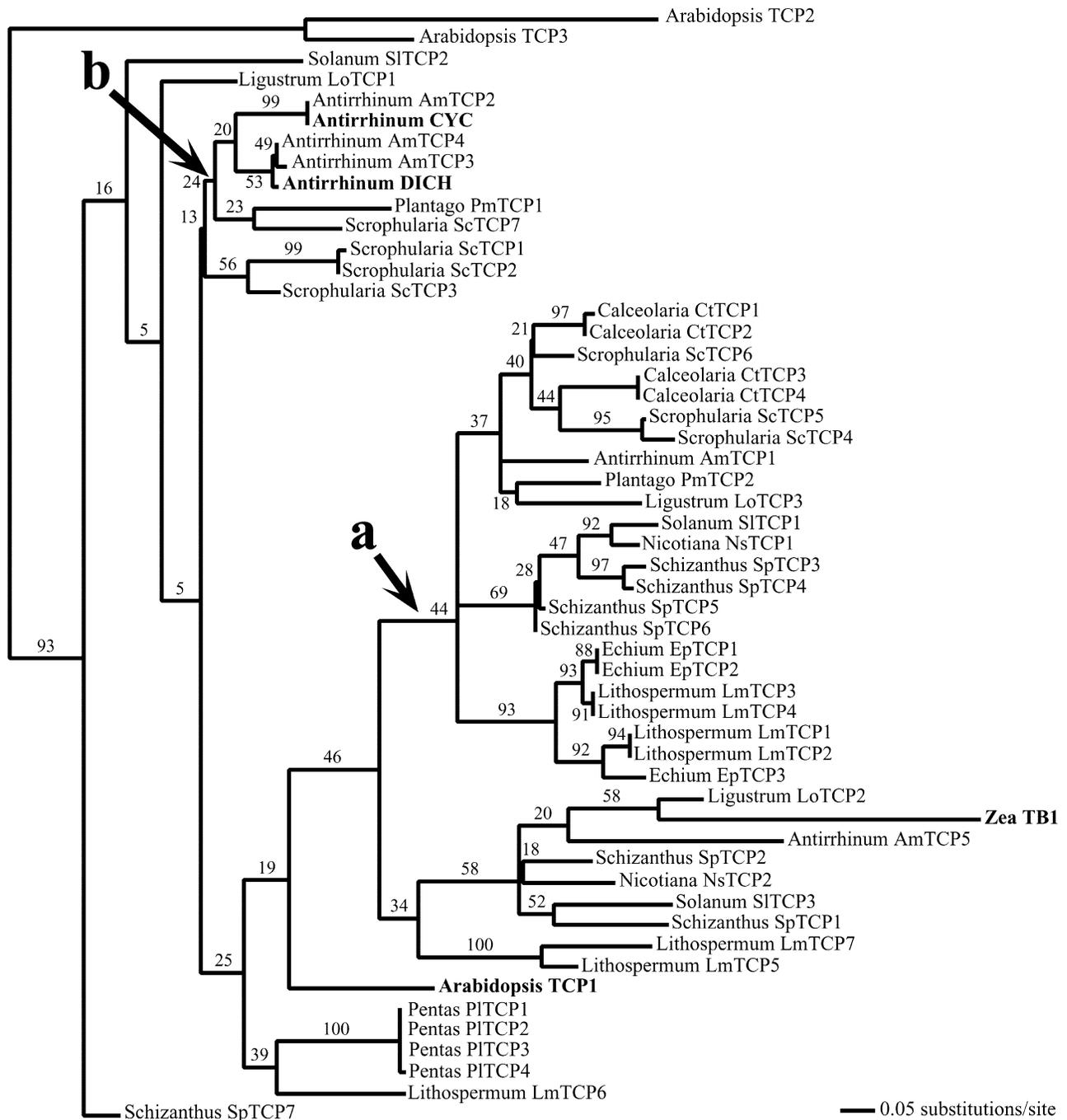


FIG. 4.—One of 10 equally likely phylogenetic trees resulting from maximum-likelihood analysis of TCP gene sequences. Sequence variant names follow generic names at tree tips. Bootstrap values for clades present in the strict consensus are shown at nodes. Bold sequence names indicate previously characterized genes. Clades where gene lineage-specific amplification bias was observed are indicated by (a) and (b). Clade (b) includes sequences from *Plantago* and *Scrophularia* which may be orthologs of *Antirrhinum cyc*.

the genomic TCP sequence diversity that could be amplified using the available primer pairs (fig. 3c). However, it is possible that using additional PCR primer pairs would increase the number of sequence variants amplified. To address this, we asked whether there was evidence for gene lineage-specific amplification bias using the primer pairs available in this study. By mapping primer pair utilization onto the ML tree, gene lineage-specific amplification bias was found when using the *cycA2*/

Rdom.1C primer combination (fig. 4). Cochran's Q test was used to determine whether some variants were more likely to be recovered than others, given random sampling of the clone pool for each taxon. A significant difference in the frequency of recovered sequence variants was found for four of the 11 taxa examined (table 4). Results (not shown) from a broader survey of *cyc/tb1* subfamily diversity suggest that as many as 150 TCP sequences may need to be recovered from a clone pool to be certain

**Table 4**  
**Results of Cochran's Q Test for Significant Differences in the Frequency of Sequence Variants in the Clone Pools**

Taxon name	P-value
<i>Antirrhinum majus</i>	0.0282
<i>Calceolaria tenella</i>	0.2153
<i>Echium plantagineum</i>	0.0686
<i>Ligustrum ovalifolium</i>	0.7788
<i>Lithospermum multiflorum</i>	$P \ll 0.0001^*$
<i>Nicotiana sylvestris</i>	0.5637
<i>Pentas lanceolata</i>	0.343
<i>Plantago major</i>	0.0047*
<i>Schizanthus aff. pinnatus</i>	$p \ll 0.0001^*$
<i>Scrophularia californica</i>	0.9856
<i>Solanum lycopersicum</i>	0.001*

NOTE.—Asterisk (\*) indicates significant at  $\alpha$  (0.05) using sequential Bonferroni correction (Rice 1989).

(binomial probability = 0.95) that all sequence variants are found.

#### Cladistic Analysis

ML analysis of the DNA data set (154 bp, EMBL ALIGN\_000466) resulted in 10 equally likely trees with a  $-\log$ -likelihood score of 3583.27836 (fig. 4). The ML tree shows strong bootstrap support at terminal branches for sister relationships among sequence variants found within individual taxa. These sequence variants are likely to be alleles at a single locus from heterozygous individuals or recently duplicated loci. Support is weaker for most clades that contain sequences from more than one taxon (i.e., those clades that may define orthologous genes).

Cladistic analysis does not provide any evidence that orthologs of *cyc* are present in the Solanaceae or Boraginaceae. All *cyc/tb1* sequences isolated from these families were found to be more closely related to *tb1* than to *cyc*, or near the base of the tree. This suggests that the mechanism by which bilaterally symmetrical flowers evolved in *Schizanthus* (Solanaceae) and *Echium* (Boraginaceae) was different from the *cyc*-mediated developmental mechanism common to bilaterally symmetrical Lamiales. Within the Lamiales, the ML tree does not indicate the presence of orthologs of *cyc* or *dich* in the basal Lamiales taxa *Ligustrum* (Oleaceae) and *Calceolaria* (Calceolariaceae), suggesting that *cyc*-like genes controlling bilateral symmetry arose after the divergence of *Ligustrum* and *Calceolaria* from the rest of the Lamiales.

#### Network Analysis

Protein genetic distances from the amino acid alignment (51 amino acids, EMBL ALIGN\_000467) were used to generate a minimum spanning network and a reticulogram to display TCP domain functional diversity (fig. 5). Within the minimum spanning network, no *cyc/tb1* sequences from Solanaceae or Boraginaceae occurred within the subnetwork of putative *cyc* complements (see Discussion for method used to define subnetwork), suggesting that a *cyc*-like function is not present in these taxa. This finding is in agreement with the cladistic analysis and is consistent with the hypothesis that the

genetic mechanism underlying bilateral flower symmetry in the Solanaceae and Boraginaceae is different from Lamiales. In contrast to the minimum spanning network, the reticulogram revealed a putative *cyc* complement in *Schizanthus* (*SpTCP7*).

No putative complements of *cyc* were identified in *Calceolaria* by either network approach. The *Ligustrum* sequence *LoTCP1* occurred within the subnetwork of putative functional analogs of *cyc* but was not found in the clade containing putative orthologs of *cyc* (fig. 4). This implies that a *cyc*-like function is present in the genome of *Ligustrum* and raises the possibility that radial flower symmetry in that taxon may be due to a difference in expression pattern, rather than gene content (as suggested by cladistic analysis). Thus, with respect to the hypothesis that the *cyc* gene lineage arose after the divergence of the basal Lamiales taxa *Ligustrum* and *Calceolaria*, the network and cladistic analyses are in conflict.

#### Gene Duplication

Between 24 and 39 duplications were necessary to reconcile the reconstructed gene genealogy with the established organismal phylogeny. A range of possible duplications is presented because the actual number is dependent on the tree topology used and whether sister sequences from the same taxon were defined as alleles at one locus (meaning a duplication event is not required to explain the observed relationship) or as distinct loci.

#### Discussion

##### Assessment of Methodology

In order for two genes from different taxa to be defined as orthologous members of a gene family, it must be determined that there are no other, as yet unidentified, candidates. The only rigorous means for making such an assertion is to establish that all candidates have been identified. Analysis of primer pair utilization suggested that screening more colonies and using more degenerate primer pairs may have resulted in the identification of additional sequence variants. Nevertheless, the procedure appears to be capable of producing a representative sampling of gene family members from a taxonomically diverse selection of species (fig. 3). By coupling this procedure to replicated taxon sampling for ancestral and derived traits, the experiment-wide probability of not finding a particular gene lineage should be minimized.

##### Cladistic-Historical Analysis

Few of the clades identified in the ML tree (fig. 4) exhibit a completely accurate reconstruction of expected organismal relationships, and branch support is generally weak. Accurate reconstruction of correct organismal relationships within and among angiosperm families (a necessity for identifying orthologs) should not be expected when only 154 bp of *cyc/tb1* sequence data are available for comparison. Furthermore, because the *cyc/tb1* subfamily is old, dating at least to the time of divergence between monocots and eudicots, gene dupli-

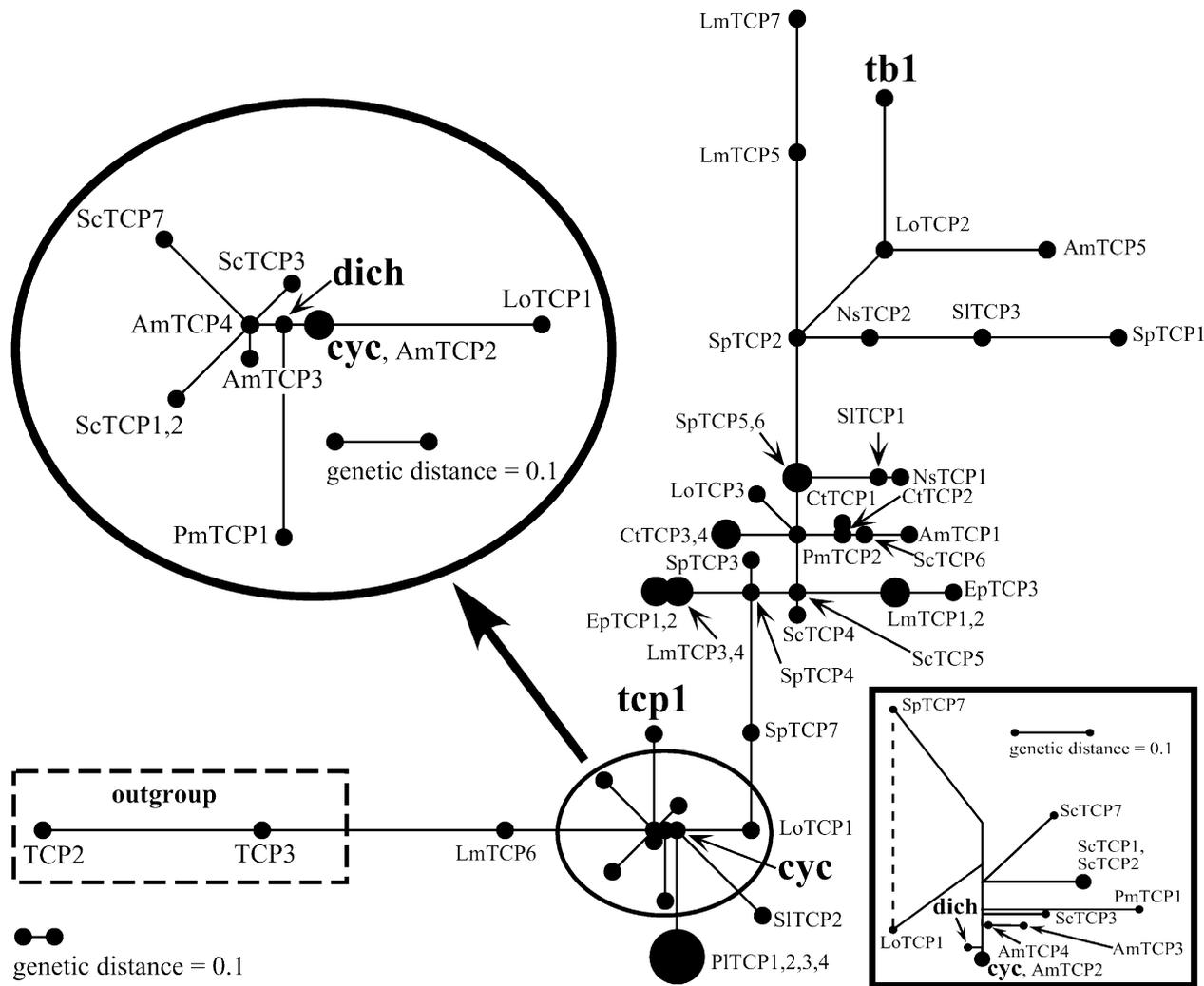


FIG. 5.—Minimum spanning network of amino acid sequence genetic distances from the TCP domain of genes isolated from Asteridae exhibiting differences in flower symmetry. The size of the filled circles at nodes is proportional to the number of sequences belonging to the node. Edge lengths are proportional to inferred protein distances. The subnetwork of sequences hypothesized to contain functional analogs of *cyc* is circled. The equivalent subnetwork from the reticulogram is shown in the box, with zero-length terminal branches collapsed. A single reticulation, inferred using the Q1 criterion (Legendre and Makarenkov 2002), is shown as a dashed line.

cation events may be ancient, and reconstruction of correct relationships between paralogous lineages also should not be expected.

If, however, it can be accepted that the reconstructed gene trees reflect some measure of historical accuracy, rapid gene duplication may be posited to have occurred during the evolution of the *cyc/tb1* subfamily. Duplication and deletion events may occur in a pattern such that cladistic analysis will result in an erroneous assertion that genes are orthologs. It might be supposed that the greatest strength of using cladistic methods to infer the genetic mechanism underlying morphological change would be for cases in which changes in gene content resulting from duplication or deletion are the causal factor. However, because it is difficult to prove that a gene is not present in a genome and because cladistic analysis may not be useful for determining orthologous relationships in the face of gene duplication and deletion, a change in gene content may be impossible to ascertain. Even if whole genome

sequences were available, fundamental limitations of using cladistic analysis to identify orthologs when gene duplication and extinction have occurred may render the hypothesis that a change in gene content is responsible for an evolutionary change in morphology nonfalsifiable.

#### Network-Functional Analysis

In cases where the use of cladistic analysis for ortholog identification is compromised by rapid gene duplication and extinction, and/or the lack of alignable or informative DNA sequence characters, another approach, focused on clustering genes with similar functional attributes may be pursued. If the degree of functional diversity of members of a multigene family can be accurately portrayed for all sampled taxa, appropriate experiments (such as reciprocal transformations or expression pattern analyses) to determine causal relationships between genetic differences and morphological variation

could be precisely defined and undertaken. In what follows, we discuss the use of networks to describe functional diversity within the *cyc/tbl* subfamily, and to generate testable hypotheses regarding the genetic mechanisms underlying floral symmetry differences in the Asteridae.

Given that the TCP domain determines DNA-binding and dimerization specificity, protein genetic distances derived from an alignment of the TCP domain may be predictive of relative differences in DNA-binding and dimerization specificity. To the extent that DNA-binding and dimerization specificities can be used to predict the regulatory function of a transcription factor, protein-genetic distances may be used to characterize functional diversity and identify putative functional analogs. As a caveat, using protein-genetic distances of regulatory regions to describe putative functional diversity is only a first approximation of true functional diversity. It is conceivable that differences in one or a few specific amino acids could have a dramatic impact on regulatory function, whereas numerous differences at nonessential sites may have no effect on the ability of two genes to complement. Therefore, protein-genetic distances are used here as a heuristic, in lieu of explicit experimental data regarding *in vivo* protein function.

#### Calibrating the Network

TCP domain sequences that are potential functional analogs may be defined as a subnetwork of the complete network based on the length of the edges connecting them. Before defining such a subnetwork, the distance at which complementation may be hypothesized to occur between the candidate gene and putative functional analogs in other taxa must be estimated. Any genes connected to the candidate gene by edges in the network shorter than this distance may be considered candidate functional analogs. Genes belonging to such a subnetwork would be predicted to have identical regulatory roles in reciprocal transformation experiments between species, provided that their expression pattern could be manipulated to mimic that of the endogenous gene.

Experimental data may be used to define a maximum genetic distance beyond which complementation would not be expected. In the minimum spanning network in figure 5, the *Arabidopsis* gene *tcp1* is connected to *AmTCP4*, a likely *dich* variant. Despite its asymmetric expression pattern (Cubas, Coen, and Zapater 2001), *tcp1* does not function in the establishment of bilateral symmetry in *Arabidopsis* (as its nearest neighbor, *AmTCP4*, likely does in *Antirrhinum*), because *Arabidopsis* has radially symmetrical flowers. Thus, the edge-length value between *AmTCP4* and *tcp1* (minimum spanning network = 0.308, reticulogram = 0.354) may be defined as a maximum distance (the "cutoff" value), beyond which complementation would not be expected to occur. An outgroup sequence should be used to avoid excluding functionally analogous in-group sequences that have accumulated substantial neutral divergence over time.

Once a reasonable edge-length value beyond which complementation would not be expected is defined, the

subnetwork of putative regulatory complements can be extracted from the larger network. An initial subnetwork may be defined by including all nodes that are connected to the candidate gene (*cyc*) with edge-length values less than the cutoff value. Assuming that the initial subnetwork contains complements of *cyc*, the final subnetwork must be extended to include any additional nodes connected to the subnetwork by values less than the cutoff. The resulting subnetwork is speculative in that the degree to which it correctly identifies *cyc* complements is dependent on the accuracy of the distance defined as noncomplementary. In the current example, the noncomplementary distance may be an overestimate because it was based on distantly related *Arabidopsis*; thus, the subnetworks extracted here may include sequences representing genes that are not complementary to *cyc*. The estimate of the maximum distance of complementation could be improved by isolating *cyc/tbl* sequences from an actinomorphic outgroup species that is more closely related to the Asteridae than *Arabidopsis*. The result of this procedure may be that sequences from all taxa will not be included in the subnetwork. This is appropriate because in some taxa, there may be no functional analog, and in others, it may not have been isolated yet.

#### Interpreting the Subnetwork

The use of an amino acid distance network provides a simple means for quickly circumscribing neighborhoods (subnetworks) of genes with hypothetically identical regulatory capabilities using the small amount of alignable sequence available for many transcription factor gene families. In some cases, genes that appear to be orthologs based on cladistic analysis may not belong to a subnetwork of putative complements. In these instances, divergence in regulatory function is implicated, and could be experimentally examined by reciprocal transformation studies. In other cases, taxa that differ in the morphological trait under scrutiny may contain sequences that are found within a subnetwork of putative functional analogs. In these cases, a difference in expression pattern, rather than gene content or regulatory function, is implicated as the likely causal factor for the morphological difference.

#### Origin of Bilateral Symmetry in the Lamiales

Within the Lamiales, bilateral floral symmetry has been inferred to have arisen on a single occasion, with several subsequent reversions to radial symmetry (Ree and Donoghue 1999; Endress 2001). A number of studies have hypothesized that a difference in the presence of *cyc* orthologs may be responsible for these evolutionary changes (Baum 1998; Donoghue, Ree, and Baum 1998; Reeves and Olmstead 1998; Citerne, Möller, and Cronk 2000). In this study, we have attempted to test this hypothesis by sampling *cyc/tbl* genes from derived taxa in the Lamiales, which exhibit bilateral floral symmetry, as well as from basal taxa, which retain the ancestral actinomorphic floral structure.

Results of the cladistic analysis suggest that the *cyc*

gene lineage controlling bilateral symmetry in Lamiales arose subsequent to the divergence of *Ligustrum* and *Calceolaria*. Whereas *Ligustrum* flowers are actinomorphic throughout development, the flowers of *Calceolaria* are zygomorphic at maturity. Early flower development in *Calceolaria* does not show the obvious dorsal/ventral asymmetry that is associated with *cyc* gene expression (Endress 1999). It is possible that floral zygomorphy in *Calceolaria* is not controlled by a *cyc*-like gene and that whatever mechanism is responsible acts at a later time in development. However, it should be noted that the organization of *Calceolaria* flowers is quite different from the rest of the zygomorphic Lamiales. *Calceolaria* has four sepals and petals and two stamens, compared with five sepals and petals and four or five stamens in most Lamiales. This raises the possibility that the action of a *cyc* ortholog may not manifest itself in the same way in *Calceolaria* as it does in *Antirrhinum*, because the field for gene expression is different. In this case, the possibility that the *cyc* ortholog from *Calceolaria* was not recovered cannot be ruled out.

The discovery of a putative *cyc*-like function in *Ligustrum* (fig. 5, *LoTCP1*) suggests that expression pattern differences may be responsible for the difference in flower symmetry between radially symmetrical *Ligustrum* and bilaterally symmetrical Lamiales. However, *LoTCP1* was not found in the clade containing putative *cyc* orthologs in the cladistic analysis (fig. 4). Given the difficulties in reconstructing genealogical relationships for data sets such as this, it is possible that cladistic analysis failed to recover the accurate position of *LoTCP1* in the tree. Regardless of whether *LoTCP1* is the *Ligustrum* ortholog of *cyc*, it is likely that *LoTCP1* is not utilized in the same manner as *cyc* in vivo because *Ligustrum* flowers are actinomorphic.

*Plantago* has actinomorphic flowers that were derived by reversion from zygomorphic ancestors (Reeves and Olmstead 1998). The edge length between *Plantago PmTCP1* and *dich* is among the longest of the subnetworks of putative functional analogs. Given this, and the aforementioned conservative nature of the noncomplementary edge length value used, *PmTCP1* may not be able to complement *cyc* even though cladistic analysis suggests it is orthologous. This implies that the mechanism underlying the reversion to radial symmetry in *Plantago* may be divergence in regulatory function rather than modification of gene expression pattern or gene content. This hypothesis could potentially be tested by transformation of *Plantago* with *cyc* under the control of the *PmTCP1* promoter. If the DNA-binding specificity of *PmTCP1* has changed but the expression pattern has remained the same since divergence from a bilaterally symmetrical ancestor, one might expect a bilaterally symmetrical *Plantago* flower to develop in the transgenic individual.

Figure 4 suggests *Scrophularia ScTCP7* as a candidate ortholog of *cyc*. Within the minimum spanning network, the edge connecting *ScTCP7* to the subnetwork is among the longest, indicating substantial divergence from *cyc*. For *Scrophularia ScTCP3*, however, the edge length to *cyc* is approximately half that of *ScTCP7*. This suggests

that the sequence most likely to complement *cyc* is not the predicted ortholog of *cyc*. It is possible that this finding is due to a failure of cladistic analysis. However, if the gene genealogy in figure 4 is accepted as true, and if *ScTCP3* is, in fact, a *Scrophularia* complement of *cyc*, then the hypothesis that the *ScTCP3* paralog was recruited to exert the homologous regulatory function of *Antirrhinum cyc* in *Scrophularia* becomes viable. It is noteworthy that although its corolla is strongly zygomorphic, *Scrophularia* belongs to a clade (Selagineae and Manuleae plus Scrophularieae [*sensu* Olmstead et al. 2001]) where actinomorphic or near-actinomorphic corolla symmetry is the rule (Endress 1999). Because strong floral zygomorphy in *Scrophularia* likely arose secondarily from a near-actinomorphic ancestor, it should not be surprising if the regulatory genes controlling the trait are found to be different from typical Lamiales.

#### Origin of Bilateral Symmetry in the Solanaceae

Within the Solanaceae, the majority of species exhibit radially symmetrical flowers; however, zygomorphy has evolved on at least three independent occasions (Olmstead and Palmer 1992). Results of the cladistic and the minimum spanning network analyses suggest that the genetic mechanism underlying bilateral symmetry in the Solanaceae is distinct from the *cyc*-mediated mechanism proposed for Lamiales. In conflict with this contention, the reticulogram suggests that a *cyc* complement may be present in *Schizanthus* (fig. 5). Given that the value for the maximum distance at which complementation might occur is likely to be an overestimate, the importance of this contradiction becomes unclear. The genetic distance between *SpTCP7* and the most similar putative *cyc* complement is only 0.002 distance units less than the maximum distance allowed within the reticulogram.

Although bilateral symmetry in *Schizanthus* may not be controlled by an ortholog of *cyc*, other TCP genes, such as *SpTCP7*, may have been recruited to fill that role. Given the example of *tcp1* in *Arabidopsis*, which exhibits dorsal/ventral asymmetry in expression pattern during early floral development (Cubas, Coen, and Zapater 2001) but does not cause bilateral flower symmetry in that taxon, it is likely that TCP genes with expression patterns similar to *cyc* will be found in Solanaceae. Such findings should be interpreted with care. Davidson (1997) has cautioned that, because recruitment may be a common evolutionary process, the observation of similarity in expression pattern of related genes in distantly related taxa cannot be taken as an indication that the processes those genes regulate are homologous. Related genes in different taxa may be proved to exert a homologous developmental function only after demonstrating that they are controlled by orthologous upstream regulators and that they regulate the expression of the same downstream genes (Davidson 1997).

#### Origin of Bilateral Symmetry in the Boraginaceae

Results of cladistic and network analyses suggest that the genetic mechanism by which zygomorphy has evolved

in *Echium* is different from that in the Lamiales. This is not surprising since dorsal/ventral asymmetry in *Echium* flowers differs from the typical Lamiales pattern. In *Echium*, bilateral symmetry in mature flowers is achieved during development by dorsal corolla enlargement (or ventral retardation), rather than the dorsal retardation caused by asymmetric expression of the *cyc* gene in *Antirrhinum*.

## Conclusions

Analyses of *cyc/tbl* subfamily sequences suggest that the three independent evolutionary origins of bilateral symmetry examined in the Asteridae were caused by the evolution of at least two distinct (nonorthologous) genetic mechanisms. Results of the cladistic analysis cannot falsify the hypothesis that the *cyc* gene lineage is unique to Lamiales, suggesting that the origin of bilateral symmetry in the Lamiales may be due to a change in gene content. Network analyses of TCP domain amino acid sequences could not concretely identify any sequences from Solanaceae or Boraginaceae that appear capable of controlling the homologous developmental processes that, under the control of *cyc* orthologs, lead to the development of bilaterally symmetrical flowers in Lamiales. Taken together, these findings suggest that gene duplication, followed by functional divergence, may be the mechanism underlying the evolution of bilateral flower symmetry in the Lamiales.

## Acknowledgments

We thank Chris Richards for helpful suggestions on the manuscript. We also are grateful to Doug Ewing for assistance with greenhouse work at the University of Washington and to Phil Friedman for providing generous access to data processing time in Colorado State University computer labs. This research was funded by a University of Washington Royalty Research Fund grant to R.O.

## Literature Cited

- Albach, D. C., P. S. Soltis, D. E. Soltis, and R. G. Olmstead. 2001. Phylogenetic analysis of asterids based on sequences of four genes. *Ann. MO Bot. Gard.* **88**:163–212.
- Almeida, J., M. Rocheta, and L. Galego. 1997. Genetic control of flower shape in *Antirrhinum majus*. *Development* **124**:1387–1392.
- Atchley, W. R., and W. M. Fitch. 1997. A natural classification of the basic helix-loop-helix class of transcription factors. *Proc. Natl. Acad. Sci. USA* **94**:5172–5176.
- Bandelt, H.-J., V. Macaulay, and M. Richards. 2000. Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. *Mol. Phylogenet. Evol.* **16**:8–28.
- Baum, D. A. 1998. The evolution of plant development. *Curr. Opin. Plant Biol.* **1**:79–86.
- Bradley, R. D., and D. M. Hillis. 1997. Recombinant DNA sequences generated by PCR amplification. *Mol. Biol. Evol.* **14**:592–593.
- Chase, M. W., D. E. Soltis, R. G. Olmstead et al. (42 co-authors). 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Ann. MO Bot. Gard.* **80**:528–580.
- Citerne, H. L., M. Möller, and Q. C. B. Cronk. 2000. Diversity of *cycloidea*-like genes in Gesneriaceae in relation to floral symmetry. *Ann. Bot.* **86**:167–176.
- Coen, E. S. 1991. The role of homeotic genes in flower development and evolution. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **42**:241–279.
- Coen, E. S., and E. M. Meyerowitz. 1991. The war of the whorls: genetic interactions controlling flower development. *Nature* **353**:31–37.
- Coen, E. S., and J. M. Nugent. 1994. Evolution of flowers and inflorescences. *Development* **1994**(Suppl.):107–116.
- Cronn, R., M. Cedroni, T. Haselkorn, C. Grover, and J. F. Wendel. 2002. PCR-mediated recombination in amplification products derived from polyploid cotton. *Theor. Appl. Genet.* **104**:482–489.
- Cubas, P., E. Coen, and J. M. M. Zapater. 2001. Ancient asymmetries in the evolution of flowers. *Curr. Biol.* **11**:1050–1052.
- Cubas, P., N. Lauter, J. Doebley, and E. Coen. 1999. The TCP domain: a motif found in proteins regulating plant growth and development. *Plant J.* **18**:215–222.
- Cubas, P., C. Vincent, and E. Coen. 1999. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* **401**:157–161.
- Davidson, E. H. 1997. Insights from the echinoderms. *Nature* **389**:679–680.
- Donoghue, M. J., R. H. Ree, and D. A. Baum. 1998. Phylogeny and the evolution of flower symmetry in the Asteridae. *Trends Plant Sci.* **3**:311–317.
- Doyle, J. J. 1994. Evolution of a plant homeotic multi-gene family: toward connecting molecular systematics and molecular developmental genetics. *Syst. Biol.* **43**:307–328.
- Endress, P. K. 1999. Symmetry in flowers: diversity and evolution. *Int. J. Plant Sci.* **160**(Suppl. 6):S3–S23.
- . 2001. Evolution of floral symmetry. *Curr. Opin. Plant Biol.* **4**:86–91.
- Excoffier, L. 1993. MINSPNET. Distributed by the author. Department of Anthropology, University of Geneva, Geneva, Switzerland.
- Excoffier, L., and P. E. Smouse. 1994. Using allele frequencies and geographic subdivision to reconstruct gene trees within a species: molecular variance parsimony. *Genetics* **136**:343–359.
- Felsenstein, J. 1978. Cases in which parsimony and compatibility methods will be positively misleading. *Syst. Zool.* **27**:401–410.
- . 2001. PHYLIP (phylogeny inference package). Version 3.6a2. Distributed by the author. Department of Genetics, University of Washington, Seattle.
- Fitch, W. M. 1997. Networks and viral evolution. *J. Mol. Evol.* **44**(Suppl.):S65–S75.
- Henikoff, J. G., S. Pietrovski, C. M. McCallum, and S. Henikoff. 2000. Blocks-based methods for detecting protein homology. *Electrophoresis* **21**:1700–1706.
- Huelsensbeck, J. P., and D. M. Hillis. 1993. Success of phylogenetic methods in the four-taxon case. *Syst. Biol.* **42**:247–264.
- Huson, D. H. 1998. SplitsTree: a program for analyzing and visualizing evolutionary data. *Bioinformatics* **14**:68–73.
- Jeanmougin, F., J. D. Thompson, M. Gouy, D. G. Higgins, and T. J. Gibson. 1998. Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* **23**:403–405.
- Jones, D. T., W. R. Taylor, and J. M. Thornton. 1992. The rapid

- generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**:275–282.
- Kosugi, S., and Y. Ohashi. 1997. PCF1 and PCF2 specifically bind to *cis* elements in the rice proliferating cell nuclear antigen gene. *Plant Cell* **9**:1607–1619.
- . 2002. DNA binding and dimerization specificity and potential targets for the TCP protein family. *Plant J.* **30**:337–348.
- Legendre, P., and V. Makarenkov. 2002. Reconstruction of biogeographic and evolutionary networks using reticulograms. *Syst. Biol.* **51**:199–216.
- Luo, D., R. Carpenter, L. Copsey, C. Vincent, J. Clark, and E. Coen. 1999. Control of organ asymmetry in flowers of *Antirrhinum*. *Cell* **99**:367–376.
- Luo, D., R. Carpenter, C. Vincent, L. Copsey, and E. Coen. 1996. Origin of floral asymmetry in *Antirrhinum*. *Nature* **383**:794–799.
- Maddison, W. P. 1997. Gene trees in species trees. *Syst. Biol.* **46**:523–536.
- Makarenkov, V. 2001. T-REX: reconstructing and visualizing phylogenetic trees and reticulation networks. *Bioinformatics* **17**:664–668.
- Martin, A. P., and T. M. Burg. 2002. Perils of paralogy: using HSP70 genes for inferring organismal phylogenies. *Syst. Biol.* **51**:570–587.
- Moore, A. W., S. Barbel, L. Y. Jan, and Y. N. Jan. 2000. A genomewide survey of basic helix-loop-helix factors in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **97**:10436–10441.
- Olmstead, R. G., B. Bremer, K. M. Scott, and J. D. Palmer. 1993. A parsimony analysis of the Asteridae sensu lato based on *rbcL* sequences. *Ann. MO Bot. Gard.* **80**:700–722.
- Olmstead, R. G., C. W. DePamphilis, A. D. Wolfe, N. D. Young, W. Y. Elisons, and P. A. Reeves. 2001. Disintegration of the Scrophulariaceae. *Am. J. Bot.* **88**:348–361.
- Olmstead, R. G., K.-J. Kim, R. K. Jansen, and S. J. Wagstaff. 2000. The phylogeny of the Asteridae sensu lato based on chloroplast *ndhF* gene sequences. *Mol. Phylogenet. Evol.* **16**:96–112.
- Olmstead, R. G., and J. D. Palmer. 1992. A chloroplast DNA phylogeny of the Solanaceae: subfamilial relationships and character evolution. *Ann. MO Bot. Gard.* **79**:346–360.
- Olmstead, R. G., and P. A. Reeves. 1995. Evidence for the polyphyly of the Scrophulariaceae based on chloroplast *rbcL* and *ndhF* sequences. *Ann. MO Bot. Gard.* **82**:176–193.
- Page, R. D. M., and M. A. Charleston. 1997. From gene to organismal phylogeny: reconciled trees and the gene tree/species tree problem. *Mol. Phylogenet. Evol.* **7**:231–240.
- Posada, D., and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**:817–818.
- . 2001. Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol. Evol.* **16**:37–45.
- . 2002. The effect of recombination on the accuracy of phylogeny estimation. *J. Mol. Evol.* **54**:396–402.
- Purugganan, M. D., S. D. Rounsley, R. J. Schmidt, and M. F. Yanofsky. 1995. Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family. *Genetics* **140**:345–356.
- Ree, R. H., and M. J. Donoghue. 1999. Inferring rates of change in flower symmetry in asterid angiosperms. *Syst. Biol.* **48**:633–641.
- Reeves, P. A., and R. G. Olmstead. 1998. Evolution of novel morphological and reproductive traits in a clade containing *Antirrhinum majus* (Scrophulariaceae). *Am. J. Bot.* **85**:1047–1056.
- Rice, W. R. 1989. Analyzing tables of statistical tests. *Evolution* **43**:223–225.
- Rose, T. M., E. R. Schultz, J. G. Henikoff, S. Pietrokovski, C. M. McCallum, and S. Henikoff. 1998. Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Res.* **26**:1628–1635.
- Rosinski, J. A., and W. R. Atchley. 1999. Molecular evolution of helix-turn-helix proteins. *J. Mol. Evol.* **49**:301–309.
- Sanderson, M. J., and J. J. Doyle. 1992. Reconstruction of organismal and gene phylogenies from data on multigene families: concerted evolution, homoplasy, and confidence. *Syst. Biol.* **41**:4–17.
- Schwarz-Sommer, Z., P. Huijser, W. Nacken, H. Saedler, and H. Sommer. 1990. Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science* **250**:931–936.
- Smouse, P. E. 2000. Reticulation inside the species boundary. *J. Classif.* **17**:165–173.
- Swofford, D. L. 1999. PAUP\*: phylogenetic analysis using parsimony (\*and other methods). Version 4.0b10. Sinauer Associates, Sunderland, Mass.
- Templeton, A. R., K. A. Crandall, and C. F. Sing. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* **132**:619–633.
- Theissen, G., J. T. Kim, and H. Saedler. 1996. Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. *J. Mol. Evol.* **43**:484–516.
- Vieira, C. P., J. Vieira, and D. Charlesworth. 1999. Evolution of the *cycloidea* gene family in *Antirrhinum* and *Misopates*. *Mol. Biol. Evol.* **16**:1474–1483.

William Martin, Associate Editor

Accepted July 6, 2003