

Duplication and Adaptive Evolution of the Chalcone Synthase Genes of *Dendranthema* (Asteraceae)

Ji Yang,* Jinxia Huang,* Hongya Gu,* Yang Zhong,† and Ziheng Yang‡

*College of Life Sciences, Peking University, Beijing; †College of Life Sciences, Fudan University, Shanghai; and

‡Department of Biology, University College London

Chalcone synthase (CHS) is a key enzyme in the biosynthesis of flavonoids, which are important for the pigmentation of flowers and act as attractants to the pollinators. Genes encoding CHS constitute a multigene family in which the copy number varies among plant species and functional divergence appears to have occurred repeatedly. Plants of the *Dendranthema* genus have white, yellow, and pink flowers, exhibiting considerable variation in flower color. In this article, 18 CHS genes from six *Dendranthema* species were sequenced. Two of them were found to be pseudogenes. The functional *Dendranthema* CHS genes formed three well-supported subfamilies: SF1, SF2, and SF3. The inferred phylogeny of the CHS genes of *Dendranthema* and *Gerbera* suggests that those genes originated as a result of duplications before divergence of these two genera, and the function of *Dendranthema* CHS genes have diverged in a similar fashion to the *Gerbera* CHS genes; i.e., the genes of SF1 and SF3 code for typical CHS enzymes expressed during different stages of development, whereas the genes of SF2 code for another enzyme that is different from CHS in substrate specificity and reaction. Relative rate tests revealed that the *Dendranthema* CHS genes significantly deviated from clocklike evolution at nonsynonymous sites. Maximum likelihood analysis showed that the nonsynonymous-synonymous ($\omega = d_N/d_S$) rate ratio for the lineage ancestral to SF2 was much higher than for other lineages, with some sites having a ratio well above one. Positive selective pressure appears to have driven the divergence of SF2 from SF1 and SF3.

Introduction

Chalcone synthase (CHS) is a key enzyme in the biosynthesis of flavonoids. CHS catalyzes the condensation of three acetate residues from malonyl-CoA with *p*-coumaroyl-CoA to form naringenin chalcone. This is the initial step of the phenylpropanoid pathway that leads to the production of flavonoids. Flavonoids have many functions in plants. They are important for the pigmentation of flowers and, hence, act as attractants to the pollinators. Flavonoids also play an important role in protection against UV light and in defense against pathogens and insects (Koes, Quattrocchio, and Mol 1994). Therefore the products of this secondary metabolic pathway enable the plant to adapt more efficiently to a stressful environment.

Most of the enzymatic components of plant secondary metabolism are encoded by small families of genes that originated through gene duplications (Durbin, McCaig, and Clegg 2000). CHS is encoded by a multigene family in many plants, such as *Petunia* (Koes et al. 1989), *Ipomoea* (Durbin et al. 1995), *Gerbera* (Helariutta et al. 1995, 1996), and leguminous plants (Ryder et al. 1987; Wingerder et al. 1989; Ito et al. 1997). Durbin, McCaig, and Clegg (2000) reviewed CHS evolution in flowering plants with special reference to the genus *Ipomoea*. They pointed out that new CHS genes are recruited recurrently in flowering plants, and the rate of nucleotide substitution was frequently accelerated in new duplicate genes. There is growing evidence for repeated divergence in the CHS gene family (Tropf et al. 1994; Helariutta et al. 1996). One example is the func-

tional shift from CHS to stilbene synthase (SS), which appears to have occurred repeatedly in plants (Tropf et al. 1994). SSs are similar to CHS in that they are polyketide synthases, but the final products of the synthetic pathway in which SSs are involved are stilbene phytoalexins rather than flavonoids (Tropf et al. 1995). The SSs are thought to have evolved independently from CHS several times over the course of plant evolution (Tropf et al. 1994), and only a small number of amino acid changes are necessary to convert CHS to SS function (Tropf et al. 1995). A second example of functional divergence of the CHS gene is found in *Gerbera hybrida* (Asteraceae). Helariutta et al. (1995) analyzed the enzymatic properties of CHS-like genes and showed that their properties are distinct from both CHS and SS. Substrate testing showed that the novel CHS gene was unable to use 4-coumaroyl-CoA as a substrate but was able to use benzoyl-CoA. A novel product is produced by the enzymatic reaction, although the role of this product in plants is unknown.

Comparison of CHS gene sequences from different species revealed that the CHS gene is structurally conserved, and most of the CHS genes contain two exons and one intron. The position of the intron is conserved. Exon 1 usually encodes 37–64 amino acid residues and exon 2 encodes about 340 residues. The latter is more conserved than the former in length and encodes almost all the active sites. The length of the intron varies significantly in different species, ranging from less than 100 bp to several kilobases.

Plants of *Dendranthema* exhibit considerable variation in flower color. Some of them are famous ornamental plants, e.g., florists *dendranthema* (formerly known as florists *chrysanthemum*). The diversity in flower color in *Ipomoea* is almost certainly due to differences in either the structure or the regulatory genes of the flavonoid biosynthetic pathway (Durbin et al.

Key words: duplication, adaptive evolution, chalcone synthase, gene family, *Dendranthema*.

Address for correspondence and reprints: Ziheng Yang, Department of Biology, University College London, Darwin Building, Gower Street, London WC1E 6BT, U.K. E-mail: z.yang@ucl.ac.uk.

Mol. Biol. Evol. 19(10):1752–1759. 2002

© 2002 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

1995). In chrysanthemum cultivar MoneyMaker, flower color was altered by modifications to the CHS gene (Courtney-Gutterson et al. 1994). It is thus interesting to examine whether the diversity in flower color correlates with the diversity of CHS gene in *Dendranthema*. In this article the *Dendranthema* CHS genes were cloned and sequenced from six species, four of which (*D. indicum*, *D. indicum* var. *aromaticum*, *D. lavandulifolium*, and *D. nankingense*) have yellow flowers, whereas *D. chanelii* has pink flowers and *D. vestitum* has white flowers. By performing phylogenetic and evolutionary analyses, we hope to characterize the organization and evolution of the CHS gene family in *Dendranthema*.

Gene duplication is considered to be a major mechanism for evolutionary innovations and functional divergence (Ohno 1970; Ohta 1993; Force et al. 1999). There has been considerable debate as to whether rapid evolution in gene families is caused by positive Darwinian selection after gene duplication (Ohta 1993) or by relaxation of the functional constraints in redundant genes (Kimura 1983, pp. 104–113). The CHS gene family provides an interesting case for testing those hypotheses. To examine the differences of selective pressures among evolutionary lineages, maximum likelihood (ML) models of codon substitution were used to analyze the functional sequences in the *Dendranthema* CHS gene family (Yang 1998; Yang and Nielsen 1998). These models used the nonsynonymous-synonymous rate ratio ($\omega = d_N/d_S$) as an indicator of selective pressure on the protein. An ω greater than 1 means that nonsynonymous mutations were fixed at a higher rate than synonymous mutations and that protein evolution is driven by positive selection. Our analysis provides evidence for positive selection driving functional divergence after gene duplication in the *Dendranthema* CHS gene family.

Materials and Methods

Plant materials used in this study are shown in table 1. Silica-dried leaves were collected for DNA isolation. PCR amplifications were performed using two primers 5'CCK TCH YTG GAY GCN MGR CAR GAC3' and 5'GG BCC RAA NCC RAA NAR MAC ACC3', as designed by Wang et al. (2000). Ambiguous nucleotides are according to the definition of Nomenclature Committee of the International Union of Biochemistry (NC-IUB 1985). The two primers correspond to the conserved regions of CHS, PSLDARQD, and RVLFGFGP and span nearly the entire coding region of exon 2. The PCR products were separated by electrophoresis, and the CHS gene fragment, about 860 bp in length, was purified from the agarose gel and cloned into a pBluescript Stratagene Chemistry vector and then sequenced using a 373A automatic DNA sequencer (Applied Biosystems, Inc). Twenty-four clones were randomly chosen and sequenced from each of the six species of *Dendranthema*. The clones obtained were denoted by the abbreviation of their source species names (the first letter of generic name and the first letter of specific epithet) and the clone number. For example, DICHS3 is clone number 3 of the CHS gene from *D. indicum*. Additional CHS sequences

Table 1
Sources of CHS Sequences

Taxon	Sequence	GenBank Accession No.	
Asteraceae			
<i>Dendranthema indicum</i>	DICHS3	AF511459 ^a	
	DICHS4	AF511460 ^a	
	DICHS24	AF511461 ^a	
	DICHS39	AF511462 ^a	
	<i>D. indicum</i> var. <i>aromaticum</i> . . .	DIACHS15	AF511463 ^a
		<i>D. lavandulifolium</i>	DLCHS1
	DLCHS8		AF511465 ^a
	<i>D. nankingense</i>	DLCHS14	AF511466 ^a
		DLCHS31	AF511467 ^a
		DNCHS16	AF511468 ^a
DNCHS17		AF511469 ^a	
<i>D. chanelii</i>	DNCHS18	AF511470 ^a	
	DCCHS1	AF511471 ^a	
<i>D. vestitum</i>	DCCHS6	AF511472 ^a	
	DVCHS34	AF511473 ^a	
<i>Gerbera hybrida</i>	DVCHS43	AF511474 ^a	
	GHCHS1	Z38096	
	GHCHS2	Z38097	
<i>Gerbera</i> sp.	GHCHS3	Z38098	
	GSPCHS	X91339	
<i>Callistephus chinensis</i>	CCCHSMR	Z67988	
Solanaceae			
<i>Solanum tuberosum</i>	STU47738	U47738	
	STU47739	U47739	
	STU47740	U47740	
<i>Lycopersicon esculentum</i>	LETCHS1	X55194	
	LETCHS2	X55195	
<i>Nicotiana tabacum</i>	AF311783	AF311783	
<i>Petunia</i> × <i>hybrida</i>	PHCHSA	X14591	
	PHCHSB	X14592	
	PHCHSD	X14593	
	PHCHSJ	X14597	
	PHCHSR	X04080	
Convolvulaceae			
<i>Ipomoea purpurea</i>	IpurA	U15946	
	IpurB	U15947	
	IpurC	U15949	
	IpurD	AB001826	
	IpurE	AB001827	
<i>Ipomoea nil</i>	InilA	U15943	
	InilB	U15944	
	InilD	AB001818	
	InilE	AB001819	
Lamiaceae			
<i>Scutellaria baicalensis</i>	AF035622	AF035622	
	AB008748	AB008748	
<i>Perilla frutescens</i>	PFAB2815	AB002815	
	PFAB2582	AB002582	
Scrophulariaceae			
<i>Antirrhinum majus</i>	AMCHS	X03710	
<i>Digitalis lanata</i>	DLAJ2526	AJ002526	

^a Sequences determined in this study.

available for other genera of Asteraceae and from closely related families in Asteridae were retrieved from GenBank and used in the phylogenetic analysis. Some of the sequences were very short and thus not used. To reduce the computation for likelihood ratio tests (LRTs) of adaptive evolution in Asteraceae CHS genes, we did not use all available outgroup sequences; instead we selected one sequence from each of the closely related families of Asteridae CHS genes as outgroups.

Sequences were first aligned at the amino acid level using Clustal X (Thompson et al. 1997). Then nucleic acid sequences were aligned according to the protein alignments. The MEGA2 program (Kumar et al. 2001) was used to estimate the number of nucleotide substitutions per synonymous and nonsynonymous site between groups of sequences (Nei and Gojobori 1986). Neighbor-joining (NJ; Saitou and Nei 1987), maximum parsimony (MP), and ML (Felsenstein 1981) methods, implemented in PAUP* 4.0 (Swofford 1998), were used for phylogenetic analysis. NJ analyses were performed using the JC69 (Jukes and Cantor 1969), K80 (Kimura 1980), and F84 (Felsenstein 1981) distance measures to examine their effects on topological stability. Heuristic tree search under parsimony was conducted using the TBR (tree-bisection-reconnection) swapping algorithm. The HKY85 model of sequence evolution (Hasegawa, Kishino, and Yano 1985; Yang 1994) was employed for ML analysis using a NJ starting tree. To reduce the effects of compositional bias on phylogenetic reconstruction, the third codon positions were excluded, although including them produced very similar phylogenies. The robustness of the tree topology was assessed by bootstrap analysis, with 1,000 resampling replicates for the MP and NJ methods and 100 replicates for the ML method.

To test for possible rate changes following gene duplication, the method of Muse and Gaut (1994), implemented in Hyphy (Pond 2001), and the method of Li and Bousquet (1992), implemented in RRTree (Robinson-Rechavi and Huchon 2000), were used for relative rate tests, with PHCHSA (*Petunia* × *hybrida* CHSA) used as the reference sequence. The method of Muse and Gaut is a LRT of rate constancy between two lineages with reference to a third outgroup lineage. The method of Li and Bousquet is a distance-based method that compares substitution rates between monophyletic groups of sequences. The test was applied to synonymous and nonsynonymous rates separately, with the method of Li (1993) being used for rate estimation.

Codon substitution models implemented in the codeml program in the PAML package (Yang 1997) were used to analyze changes in selective pressure during functional divergences of the *Dendranthema* CHS gene family. Two kinds of codon-substitution models are used. The “branch” models allowed the ω ratio to vary among lineages and were used to construct LRTs to examine whether the ω ratio along lineages after gene duplication was higher than along other lineages. Those models average synonymous and nonsynonymous rates over all sites in the sequence. The “branch-site” model accounts for variation in selective pressure among sites and is used to test for positive selection along the branches of interest which affects only a few amino acid sites (Yang and Nielsen 2002). The model assumes four classes of sites. The first two site classes have ω_0 and ω_1 along all lineages in the phylogeny. The third and fourth site classes have ω_0 and ω_1 along all branches except a few branches of interest, which have ω_2 . When the estimate of ω_2 is greater than 1, some sites are under positive selection along the branches of interest. This

model can be compared with a “site-specific” model (M3 discrete, $K = 2$; Yang et al. 2000), that allows for two site classes with ω_0 and ω_1 only, to construct a LRT.

Results

Duplications and Sequence Divergences of *Dendranthema* CHS Genes

Twenty-four clones were chosen at random for sequencing from each of the six species of *Dendranthema*. Eighteen different fragments of CHS genes were obtained. Their GenBank accession numbers are AF511459–AF511476. Aligning those sequences revealed a 3-base and two 1-base deletions in sequence DCCHS5 (GenBank accession number AF511475) and a 37-base deletion in sequence DVCHS49 (GenBank accession number AF511476). Those two sequences were confirmed by repeated PCR reaction, cloning, and sequencing. The frameshift mutations lead to premature stop codons, indicating that DCCHS5 and DVCHS49 are nonfunctional pseudogenes. These two sequences were not used in later analysis. Alignment of the functional sequences was straightforward, and no alignment gaps need be introduced. All the functional fragments include 855 nucleotides, coding for 285 amino acid residues. Sequence identity ranged from 72.8% to 98.7% at the nucleotide level (including all three codon positions) and from 72.2% to 99.6% at the amino acid level.

The phylogenetic relationships between *Dendranthema* CHS genes and the CHS-like sequences available for Asteridae were inferred by NJ, MP, and ML analyses. NJ trees were reconstructed with several distance measures (JC69, K80, and F84), and a parsimony tree search was conducted using the TBR perturbation algorithm. The HKY85 model of sequence evolution was used for inferring the ML tree. The tree topologies produced by different methods were similar on the overall structure. Furthermore, analyzing the first and second codon positions only and analyzing all three codon positions produced similar results. Figure 1 shows the phylogenetic tree reconstructed by ML using the first and second codon positions. It highlights three important features in all inferred trees. First, the CHS sequences of *Dendranthema* formed three distinct subfamilies, designated here as SF1, SF2, and SF3 (fig. 1). Indeed the different methods produced identical relationships between the subfamilies, although the within-subfamily relationships are not stable. Second, the CHS genes of *Dendranthema* did not form species-specific clusters but instead formed subfamilies with the CHS genes from other plants. For example, sequences DICHS39, DICHS24, and DICHS3 were from the same species *D. indicum*, but they were grouped into different subfamilies. Third, the three subfamilies of the *Dendranthema* CHS gene clustered with GHCHS1, GHCHS2, and GHCHS3 of *Gerbera*, respectively. Those three genes have been studied in detail both on gene structure and function. Helariutta et al. (1995) pointed out that GHCHS1, GHCHS2, and GHCHS3 are different members of the *Gerbera* CHS gene family, and GHCHS1 and GHCHS3 code for typical CHS enzyme, whereas the

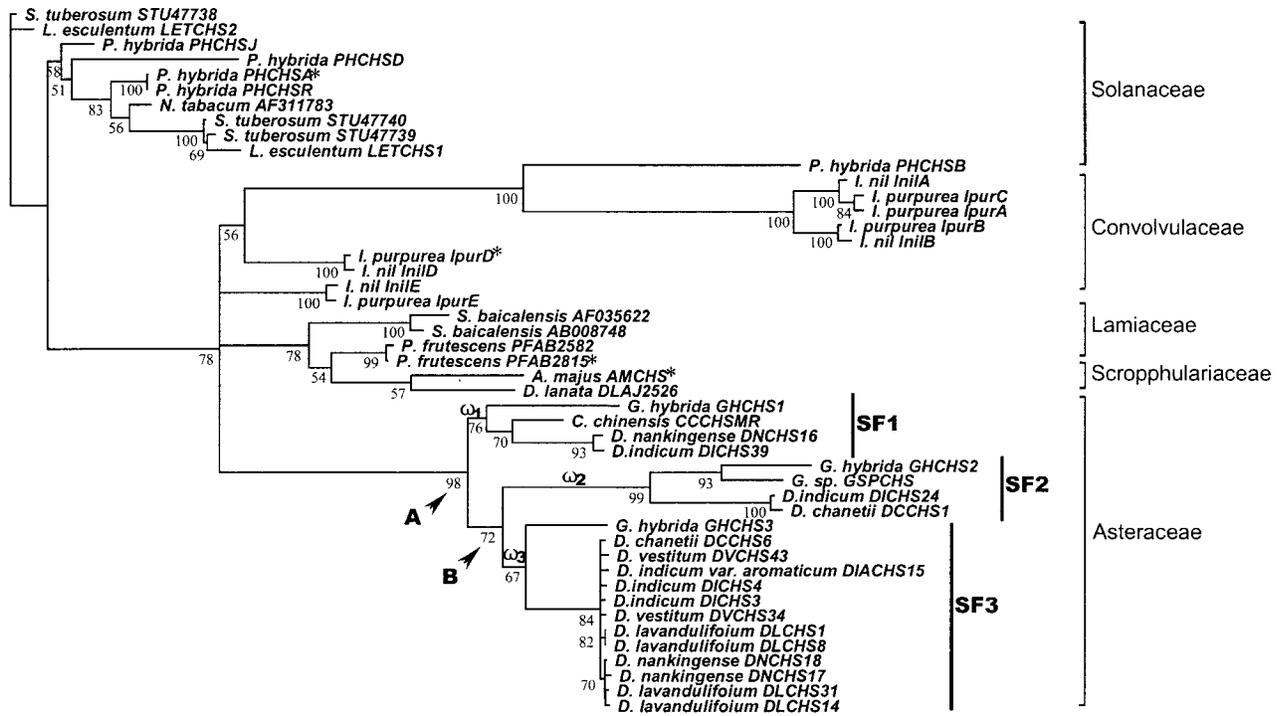


FIG. 1.—ML tree for CHS genes from *Dendranthema* and CHS-like sequences for Asteridae. Sequences both by the species name and the clone number. See table 1 for abbreviations of genus names. The HKY85 substitution model was applied to the first and second codon positions. The numbers at internal nodes are bootstrap proportions. Values under 50% are not shown. Branch lengths are proportional to the expected number of nucleotide substitutions per site. The codon-based analysis uses all sequences in the Asteraceae family and four sequences, indicated by asterisks, from the four outgroup families: Solanaceae, Convolvulaceae, Lamiaceae, and Scrophulariaceae.

GHCHS2 enzyme differs from CHSs in its substrate specificity and reaction. CHS genes in *Dendranthema* showed a similar divergence pattern and fell into three subfamilies as well. Sequences DNCHS16 and DICHS39 were homologous to GHCHS1, and sequences DICHS24 and DCCHS1 were homologous to GHCHS2, whereas the rest of the *Dendranthema* CHS sequences were more closely related to GHCHS3 (fig. 1). The phylogenetic analysis thus suggests that the CHS genes of *Dendranthema* and *Gerbera* originated from common ancestral genes, and the duplications giving rise to those ancestral genes occurred before the divergence of *Dendranthema* and *Gerbera*.

Synonymous and nonsynonymous distances between the *Dendranthema* CHS gene subfamilies were estimated using the method of Nei and Gojobori (1986) (table 2). The results show that at the synonymous sites, SF2 and SF3 are similar, consistent with the phylogenetic relationships (fig. 1). At the nonsynonymous sites SF1 and SF3 are similar, whereas SF2 is different, in-

dicating that many nonsynonymous substitutions might have occurred along the SF2 lineage.

Molecular Clocks and Rate of Evolution

To examine the rate variation between members of the CHS gene family in *Dendranthema*, we performed the relative rate test of Muse and Gaut (1994). Codon models were used to test whether two sequences followed either a nonsynonymous or a synonymous molecular clock. The outgroup sequence in all comparisons was PHCHSA. No significant rate difference was detected within each of the three subfamilies for either the synonymous or nonsynonymous rates. Similarly no difference was detected at the synonymous sites in between-subfamily comparisons. But nonsynonymous rates were significantly different between SF2 and SF1 or SF3; the smallest log-likelihood difference in all the comparisons was 20, with $P < 0.001$, $df = 1$. No significant nonsynonymous rate difference was found between SF1 and SF3. Those results were substantiated by the relative rate test of Li and Bousquet (1992), which compares evolutionary rates between groups of sequences, taking into account the covariances between sequences. The results suggest that SF2 evolved significantly faster than SF1 ($Z = 5.30$, $P < 0.0001$) and SF3 ($Z = 5.23$, $P < 0.0001$) at nonsynonymous sites. In contrast, synonymous rates were homogeneous between the three subfamilies ($P \geq 0.78$).

Table 2
Estimated Numbers of Synonymous (Below Diagonal) and Nonsynonymous (Above Diagonal) Substitutions per Site Between Subfamilies of the *Dendranthema* CHS Genes

	SF1	SF2	SF3
SF1		0.142	0.046
SF2	0.639		0.138
SF3	0.681	0.602	

Table 3
Log-likelihood Values and Parameter Estimates Under Different Models

Model	<i>p</i>	ℓ	κ	ω_0	ω_1	ω_2	ω_3
A $\omega_0 = \omega_1 = \omega_2 = \omega_3 \dots$	49	-6881.85	1.701	0.053	ω_0	ω_0	ω_0
B $\omega_0 = \omega_2 = \omega_3, \omega_1 \dots$	50	-6880.80	1.700	0.052	0.207	ω_0	ω_0
C $\omega_0 = \omega_1 = \omega_3, \omega_2 \dots$	50	-6863.81	1.700	0.047	ω_0	0.528	ω_0
D $\omega_0 = \omega_1 = \omega_2, \omega_3 \dots$	50	-6881.63	1.700	0.054	ω_0	ω_0	0.033
E $\omega_0, \omega_1 = \omega_2 = \omega_3 \dots$	50	-6868.71	1.708	0.046	0.207	ω_1	ω_1
F $\omega_0, \omega_1, \omega_2, \omega_3 \dots$	52	-6862.44	1.700	0.046	0.152	0.499	0.023
G $\omega_0 = \omega_1, \omega_2, \omega_3 \dots$	51	-6863.69	1.700	0.047	ω_0	0.543	0.033
H $\omega_0 = \omega_2, \omega_1, \omega_3 \dots$	51	-6879.33	1.697	0.053	0.199	ω_0	0.017
I $\omega_0 = \omega_3, \omega_1, \omega_2 \dots$	51	-6862.87	1.700	0.046	0.117	0.497	ω_0

NOTE.—*p*, number of parameters.

As the molecular clock assumption appears to hold at synonymous sites, we use synonymous rates to calculate rough estimates of the gene duplication times. No good fossil data are available to calibrate the clock. Instead we use the average synonymous substitution rate for plant nuclear genes at about 5×10^{-9} substitutions per site per year (Li 1997, p. 193). The average synonymous distance from SF1 to SF2 and SF3 was 0.66 substitutions per site. The average synonymous distance between SF2 and SF3 was 0.602 substitutions per site. Thus the times of duplications were estimated to be about 66 (node A in fig. 1) and 60 (node B) million years ago. The divergence time of Asteraceae was estimated to be at 30 to 60 million years ago (Cronquist 1977).

LRTs of Adaptive Evolution after Gene Duplication

To understand the mechanisms of evolutionary rate variation among lineages, we apply two kinds of likelihood rate tests based on models of codon substitution. The first analysis examines the variation of selective pressures among lineages, with the selective pressure indicated by the nonsynonymous-synonymous rate ratio ($\omega = d_N/d_S$) (Yang 1997, 1998; Yang and Nielsen 1998). The phylogeny of Asteraceae CHS genes was assumed as shown in figure 1 but only four sequences from the four outgroup families (PHCHSA, IpurD, PFAB2815, and AMCHS) were used. Three different models were used. The “one-ratio” model assumes the same ω ratio for all lineages (fig. 1; table 3). The log-likelihood value under this model was $\ell_0 = -6881.85$, with the estimate $\omega = 0.053$. The low ω ratio highlights the overwhelming role of purifying selection in this gene family. The

“free-ratios” model assumes an independent ω ratio for each branch in the tree. The likelihood value under this model was $\ell_1 = -6748.87$. Comparison of twice the log-likelihood difference, $2\delta\ell = 2(\ell_1 - \ell_0) = 2 \times (-6748.87 - (-6881.85)) = 265.96$, with the χ^2 distribution (df = 46) suggested rejection of the one-ratio model, with $P < 0.001$. The difference between the two models was significant, indicating that the ω ratios were extremely variable among lineages.

To detect changes in selective pressure after gene duplication, the “two-ratios” model was used, which assumes that the branches of interest have different ω ratios from the background ratio ω_0 (Yang 1998). Here we assigned ω_1, ω_2 , and ω_3 to the lineages ancestral to SF1, SF2, and SF3, respectively, and fitted several models with different assumptions concerning those ω ratios (table 3). Estimates of the ω ratios for all the ancestral lineages were less than 1. But the ω ratios differed among the ancestral lineages, with ω_2 for the lineage ancestral to SF2 being much higher than for other lineages. The log-likelihood values under different models were compared to test various hypotheses (table 4). The results suggest that ω_1 and ω_3 were not significantly different from the background ratio ω_0 , but ω_2 was significantly greater than ω_0 . The LRT results, combined with the results of the relative rate tests, suggest an increase in the nonsynonymous substitution rate in the lineage ancestral to SF2 after the gene duplication.

The above analysis averages the ω ratio over all codons in the gene. However, changes to a few key residues at the active sites of a protein might be sufficient to cause functional divergence, but these changes would represent too small a fraction of the overall changes to elevate d_N substantially. Thus we performed a second analysis on the same data to test for positive selection affecting individual sites along the lineage ancestral to the SF2 genes (fig. 1). The branch-site model of Yang and Nielsen (2002), is used, as implemented in the codeml program of the PAML package (Yang 1997). Parameter estimates under this model suggested that 21.8% of codons in the gene were under positive selection with $\omega_2 = 2.7$ along that lineage. Because those estimates are based on information about changes along one single branch (the branch ancestral to SF2), the exact values may not be very reliable. However, we have to trust the estimates only to the extent that the proportion is >0

Table 4
Likelihood Ratio Statistics ($2\delta\ell$) for Testing Hypotheses

Null Hypothesis Tested	Assumption Made	Models Compared	$2\delta\ell$
A' $(\omega_1 = \omega_2 = \omega_3) = \omega_0 \dots$	$\omega_1 = \omega_2 = \omega_3$	A and E	26.28*
B' $\omega_1 = \omega_0 \dots$	$\omega_2 = \omega_3 = \omega_0$	A and B	2.10
C' $\omega_1 = \omega_0 \dots$	ω_2, ω_3	G and F	2.50
D' $\omega_2 = \omega_0 \dots$	$\omega_1 = \omega_3 = \omega_0$	A and C	36.08*
E' $\omega_2 = \omega_0 \dots$	ω_1, ω_3	H and F	33.78*
F' $\omega_3 = \omega_0 \dots$	$\omega_1 = \omega_2 = \omega_0$	A and D	0.42
G' $\omega_3 = \omega_0 \dots$	ω_1, ω_2	I and F	0.86

* Significant ($P < 1\%$).

and the ratio $\omega_2 > 1$. The LRT comparing the branch-site model and the site-specific model M3 ($K = 2$) gave $2\delta\ell = 30.54$, much greater than the critical value $\chi^2_{1\%} = 9.21$ with $df = 2$, indicating that the branch-site model fits the data significantly better. The LRT thus provides significant evidence for positive selective pressure driving functional divergence of SF2 genes.

Discussion

More than 100 CHS genes have been cloned from various plants (Yamazaki et al. 2001). However, what remains unclear is how many members the CHS gene family is composed of in each species and how to distinguish different members. In this article, we randomly screened twenty-four clones for each of the six *Dendranthema* species and obtained sixteen functional sequences. Phylogenetic analysis showed that the CHS genes of *Dendranthema* formed three distinct subfamilies, SF1, SF2, and SF3, just as in the *Gerbera* genus. Because not all copies of the CHS gene were obtained from each of the six species, we cannot claim that SF1 and SF2 are not present in all those species. The phylogenetic analysis indicates that the gene duplications giving rise to the subfamilies predate the divergence between *Dendranthema* and *Gerbera*. According to Helariutta et al. (1995), sequences GHCHS1 and GHCHS3 code for typical CHS enzymes. GHCHS3 expression correlates with early synthesis of flavonols and GHCHS1 with later synthesis of both flavonols and anthocyanins. However, GHCHS2 is different. The expression pattern does not correlate with the pigmentation pattern; the amino acid sequence deviates considerably from the consensus sequence of typical CHS, and the catalytic properties are different. On the basis of the phylogenetic relationships among the CHS genes of *Dendranthema* and *Gerbera* (fig. 1), it seems reasonable to expect that the *Dendranthema* CHS genes have diverged in a similar fashion to the *Gerbera* CHS genes.

Gene duplication is often followed by an elevated evolutionary rate (Li 1985; Zhang, Rosenberg, and Nei 1998; Duda and Palumbi 1999; Schmidt, Goodman, and Grossman 1999; Bielawski and Yang 2001), which can be due to either positive Darwinian selection for functional divergence (Ohta 1993) or relaxation of selective constraints (Kimura 1983). In the former case, the requirement of the new function exerts directional selective pressure, promoting fixation of advantageous nonsynonymous mutations. In the latter case, neutral mutations are fixed at random, which, perhaps due to environmental changes, lead to a novel function in one or both copies. The two hypotheses are often difficult to distinguish; for example, an elevated ω ratio that is not >1 is compatible with both hypotheses.

In the present analysis of the *Dendranthema* and *Gerbera* CHS genes, the relative rate test suggested homogeneous synonymous rates, whereas nonsynonymous rates are different when sequences from SF2 are compared with those from SF1 or SF3. Likelihood analysis using the branch models showed that the ω ratios were highly variable among lineages of the tree, and the ω

ratio for the branch ancestral to SF2 is much higher than for other lineages (table 2). Although the estimated ω ratio is not greater than 1, this ratio is an average across all codons in the gene. A second analysis using the branch-site model accounts for heterogeneous selective pressure among sites and provided significant evidence for positive selections (with $\omega > 1$) acting at some sites along the branch ancestral to SF2. Furthermore, studies of *Gerbera* CHS genes indicate different functions of GHCHS2 from GHCHS1 and GHCHS3 (Helariutta et al. 1995), and the phylogeny of figure 1 suggests similar functional divergence of SF2 from SF1 and SF3 in *Dendranthema*. Combining those results, we conclude that positive selection is a more likely explanation for the evolution of the *Dendranthema* CHS gene family than relaxed selective constraints.

The species sampled in this article have different flower colors. Yet, no correlation can be easily found between the flower color and the CHS genes of *Dendranthema*. The yellow-flowered species, *D. indicum*, has CHS genes from all three subfamilies, whereas other yellow-flowered species, *D. indicum* var. *aromaticum*, *D. nankingense*, and *D. lavandulifolium*, have CHS genes from only some of the subfamilies (fig. 1). The white-flowered *D. vestitum* has CHS genes from SF3 only, and the pink flowered species *D. chanelii* has CHS genes from both SF2 and SF3. We note that some members of the CHS gene family may not have been sequenced in some of the species. Also, flower color might not be determined by the number of CHS family members. Furthermore, other loci may affect flower color as well. Clegg, Cummings, and Durbin (1997) examined flower color variation in *Ipomoea purpurea* and suggested that as many as five loci control floral phenotypes in *I. purpurea*. Of those, only the *A/a* locus, which encodes the CHS gene, is well characterized at the molecular level, whereas the other four are not. In *Dendranthema* very few studies have been conducted on the molecular biology of the genes of flavonoid biosynthesis that determine flower color, and we do not yet know how many genes are responsible for flower color polymorphisms of *Dendranthema* plants. More work is needed for a complete causal analysis that connects floral phenotypes to genes.

Acknowledgments

We thank two anonymous referees for comments. This study was supported by the State Key Basic Research and Development Plan of China grants (G2000046804) to J.Y., National Natural Science Foundation of China grants (39830131 and 39970051) to H.G. and J.Y., and BBSRC grants to Z.Y. Part of this work was completed during J.Y.'s visit to UCL, supported by a short-visit fellowship from the Royal Society.

LITERATURE CITED

- BIELAWSKI, J. P., and Z. YANG. 2001. Positive and negative selection in the DAZ gene family. *Mol. Biol. Evol.* **18**:523–529.

- CLEGG, M. T., M. P. CUMMINGS, and M. L. DURBIN. 1997. The evolution of plant nuclear genes. *Proc. Natl. Acad. Sci. USA* **94**:7791–7798.
- COURTNEY-GUTTERSON, N., C. NAPOLI, C. LEMIEUX, A. MORGAN, E. FIROOZABADY, and K. E. ROBISON. 1994. Modification of flower color in florist's chrysanthemum: production of a white-flowering variety through molecular genetics. *Biotechnology* **12**:268–271.
- CRONQUIST, A. 1977. The Compositae revisited. *Brittonia* **29**:137–153.
- DUDA, T. F. JR., and S. R. PALUMBI. 1999. Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. *Proc. Natl. Acad. Sci. U. S. A.* **96**:6820–6823.
- DURBIN, M. L., G. H. LEARN, G. A. HUTTLEY, and M. T. CLEGG. 1995. Evolution of the chalcone synthase gene family in the genus *Ipomoea*. *Proc. Natl. Acad. Sci. USA* **92**:3338–3342.
- DURBIN, M. L., B. MCCAIG, and M. T. CLEGG. 2000. Molecular evolution of the chalcone synthase multigene family in the morning glory genome. *Plant Mol. Biol.* **42**:79–92.
- FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17**:368–376.
- FORCE, A., M. LYNCH, F. B. PICKETT, A. AMORES, Y. L. YAN, and J. POSTLETHWAIT. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**:1531–1545.
- HASEGAWA, M., H. KISHINO, and T. YANO. 1985. Dating the human-ape split by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**:160–174.
- HELARIUTTA, Y., P. ELOMMA, M. KOTILAINEN, R. J. GRIESBACH, J. SCHRODER, and T. H. TEERI. 1995. Chalcone synthase-like genes active during corolla development are differentially expressed and encode enzymes with different catalytic properties in *Gerbera hybrida* (Asteraceae). *Plant Mol. Biol.* **28**:47–60.
- HELARIUTTA, Y., M. KOTILAINEN, P. ELOMAA, N. KALKKINEN, K. BREMER, T. H. TEERI, and V. A. ALBERT. 1996. Duplication and functional divergence in the chalcone synthase gene family of Asteraceae: evolution with substrate change and catalytic simplification. *Proc. Natl. Acad. Sci. USA* **93**:9033–9038.
- ITO, M., Y. ICHINOSE, H. KATO, T. SHIRAISHI, and T. YAMADA. 1997. Molecular evolution and functional relevance of the chalcone synthase genes of pea. *Mol. Gen. Genet.* **255**:28–37.
- JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21–123 in H. N. MUNRO, ed. *Mammalian protein metabolism*. Academic Press, New York.
- KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**:111–120.
- . 1983. *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge, U.K.
- KOES, R. E., F. QUATTROCCHIO, and J. N. M. MOL. 1994. The flavonoid biosynthetic pathway in plants: function and evolution. *BioEssays* **16**:123–132.
- KOES, R. E., C. E. SPELT, P. J. VAN DEN ELZEN, and J. N. M. MOL. 1989. Cloning and molecular characterization of the chalcone synthase multigene family of *Petunia hybrida*. *Gene* **81**:245–257.
- KUMAR, S., K. TAMURA, I. B. JAKOBSEN, and M. NEI. 2001. MEGA2: molecular evolutionary genetics analysis software. Arizona State University, Tempe, Ariz.
- LI, W.-H. 1985. Accelerated evolution following gene duplication and its implications for the neutralist-selectionist controversy. Pp. 333–352 in T. OHTA and K. AOKI, eds. *Population genetics and molecular evolution*. Japan Scientific Press, Tokyo.
- . 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J. Mol. Evol.* **36**:96–99.
- . 1997. *Molecular evolution*. Sinauer Associates, Inc., Sunderland, Mass.
- LI, P., and J. BOUSQUET. 1992. Relative-rate test for nucleotide substitutions between two lineages. *Mol. Biol. Evol.* **9**:1185–1189.
- MUSE, S. V., and B. S. GAUT. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates with application to the chloroplast genome. *Mol. Bio. Evol.* **11**:715–724.
- NEI, M., and T. GOJOBORI. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Bio. Evol.* **3**:418–426.
- NC-IUB (NOMENCLATURE COMMITTEE OF THE INTERNATIONAL UNION OF BIOCHEMISTRY). 1985. Nomenclature for incompletely specified bases in nucleic acid sequences. Recommendations. *Eur. J. Biochem.* **150**:1.
- OHNO, S. 1970. *Evolution by gene duplication*. Springer-Verlag, New York.
- OHTA, T. 1993. Pattern of nucleotide substitution in growth hormone-prolactin gene family: a paradigm for evolution by gene duplication. *Genetics* **134**:1271–1276.
- POND, S. K. 2001. Hypothesis testing using phylogenies (HYPHY). Version 0.91 beta. University of Arizona, Tucson.
- ROBINSON-RECHAVI, M., and D. HUCHON. 2000. RRTree: relative-rate tests between groups of sequences on a phylogenetic tree. *Bioinformatics* **16**:296–297.
- RYDER, T. B., S. A. HEDRICK, J. N. BELL, X. LIANG, S. D. CLOUSE, and C. J. LAMB. 1987. Organization and differential activation of a gene family encoding the plant defense enzyme chalcone synthase in *Phaseolus vulgaris*. *Mol. Gen. Genet.* **210**:219–233.
- SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
- SCHMIDT, T. R., M. GOODMAN, and L. I. GROSSMAN. 1999. Molecular evolution of the COX7A gene family in primates. *Mol. Biol. Evol.* **16**:619–626.
- SWOFFORD, D. L. 1998. PAUP*: phylogenetic analysis using parsimony (* and other methods). Version 4.0. Sinauer Associates, Sunderland, Mass.
- THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, and D. G. HIGGINS. 1997. The CLUSTAL-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**:4876–4882.
- TROPF, S., B. KARCHER, G. SCHRODER, and J. SCHRODER. 1995. Reaction mechanisms of homodimeric plant polyketide synthases (stilbene and chalcone synthase). A single active site for the condensing reaction is sufficient for synthesis of stilbenes, chalcones, and 6'-deoxychalcones. *J. Biol. Chem.* **270**:7922–7928.
- TROPF, S., T. LANZ, S. A. RENSING, J. SCHRODER, and G. SCHRODER. 1994. Evidence that stilbene synthases have developed from chalcone synthases several times in the course of evolution. *J. Mol. Evol.* **38**:610–618.
- WANG, J., L. QU, J. CHEN, H. GU, and Z. CHEN. 2000. Molecular evolution of the exon 2 of CHS genes and the possibility of its application to plant phylogenetic analysis. *Chinese Sci. Bull.* **45**:1735–1742.
- WINGENDER, R., H. ROHRIG, C. HORICKE, D. WING, and J. SCHELL. 1989. Differential regulation of soybean chalcone

- synthase genes in plant defence, symbiosis and upon environmental stimuli. *Mol. Gen. Genet.* **218**:315–322.
- YAMAZAKI, Y., D. SUH, W. SITTHITHAWORN, K. ISHIGURO, Y. KOBAYASHI, M. SHIBUYA, Y. EBIZUKA, and U. SANKAWA. 2001. Diverse chalcone synthase superfamily enzymes from the most primitive vascular plant, *Psilotum nudum*. *Planta* **214**:75–84.
- YANG, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* **39**:306–314.
- . 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**:555–556 [<http://abacus.gene.ucl.ac.uk/software/paml.html>].
- . 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* **15**:568–573.
- YANG, Z., and R. NIELSEN. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J. Mol. Evol.* **46**:409–418.
- . 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **19**:908–917.
- YANG, Z., R. NIELSEN, N. GOLDMAN, and A.-M. K. PEDERSEN. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**:431–449.
- ZHANG, J., H. F. ROSENBERG, and M. NEI. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. U. S. A.* **95**:3708–3713.

RICHARD THOMAS, reviewing editor

Accepted June 1, 2002