

## Positive selection drives adaptive diversification of the 4-coumarate: CoA ligase (4CL) gene in angiosperms

Haiyan Sun<sup>1,2,3,4</sup>, Kai Guo<sup>2,3,4</sup>, Shengqiu Feng<sup>2,3,5</sup>, Weihua Zou<sup>2,3,5</sup>, Ying Li<sup>2,3,5</sup>, Chunfen Fan<sup>2,3,5</sup> & Liangcai Peng<sup>2,3,4,5</sup>

<sup>1</sup>School of Biology and Food Engineering, Changshu Institute of Technology, Changshu 215500, China

<sup>2</sup>National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China

<sup>3</sup>Biomass and Bioenergy Research Centre, Huazhong Agricultural University, Wuhan 430070, China

<sup>4</sup>College of Life Science and Technology, Huazhong Agricultural University, Wuhan 430070, China

<sup>5</sup>College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China

### Keywords

4-Coumarate: coenzyme A ligase, diversification, evolution, phylogeny, positive selection.

### Correspondence

Liangcai Peng, National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China.

Tel: 86-27-87281765;

Fax: 86-27-87280016;

E-mail: lpeng@mail.hzau.edu.cn

### Funding Information

The National Transgenic Project (2009ZX08009-119B), the 111 Project (B08032), the 973 Specific Pre-project (2010CB134401) and the Youth Foundation of Jiangsu Province (BK20140417).

Received: 11 May 2015; Revised: 25 June 2015; Accepted: 25 June 2015

*Ecology and Evolution* 2015; 5(16): 3413–3420

doi: 10.1002/ece3.1613

## Introduction

Lignin and flavonoids are thought to play vital roles in the adaptation of plants to terrestrial environments (Rozemaa et al. 2002; Weng and Chapple 2010; Agati et al. 2013). The enzyme 4-Coumarate: CoA ligase (4CL; EC 6.2.1.12) is a key enzyme that functions in an early step of the general phenylpropanoid pathway. The protein 4CL converts 4-coumaric acid and other cinnamic acids, such as caffeic acid and ferulic acid, into the corresponding CoA thiol esters, which are then subsequently used for the biosynthesis of numerous secondary metabolites, including flavonoids, isoflavonoids, lignin,

## Abstract

Lignin and flavonoids play a vital role in the adaptation of plants to a terrestrial environment. 4-Coumarate: coenzyme A ligase (4CL) is a key enzyme of general phenylpropanoid metabolism which provides the precursors for both lignin and flavonoids biosynthesis. However, very little is known about how such essential enzymatic functions evolve and diversify. Here, we analyze 4CL sequence variation patterns in a phylogenetic framework to further identify the evolutionary forces that lead to functional divergence. The results reveal that lignin-biosynthetic 4CLs are under positive selection. The majority of the positively selected sites are located in the substrate-binding pocket and the catalytic center, indicating that nonsynonymous substitutions might contribute to the functional evolution of 4CLs for lignin biosynthesis. The evolution of 4CLs involved in flavonoid biosynthesis is constrained by purifying selection and maintains the ancestral role of the protein in response to biotic and abiotic factors. Overall, our results demonstrate that protein sequence evolution via positive selection is an important evolutionary force driving adaptive diversification in 4CL proteins in angiosperms. This diversification is associated with adaptation to a terrestrial environment.

suberins, coumarins and wall-bound phenolics (Ehrling et al. 1999; Saballos et al. 2012). The 4CL gene family is typically small. The 4CL family has 4 members in *Arabidopsis* (Hamberger and Hahlbrock 2004), 5 members in rice (Gui et al. 2011; Sun et al. 2013), and 4 members in soybean (Lindermayr et al. 2002). 4CL isoforms with different substrate specificities may direct the flow from general phenylpropanoid metabolism into the different pathways for specific end products (Souza et al. 2008).

In dicots, 4CLs can be divided into two distinct groups: class I and class II. The disruption of 4CL expression has demonstrated that class I 4CLs participate in lignin

formation, while class II 4CLs impact flavonoid metabolism (Lee et al. 1997; Hu et al. 1998; Ehlting et al. 1999; Harding et al. 2002; Nakashima et al. 2008). The remarkable functional diversity of 4CL suggests that it may be subject to positive Darwinian selection. However, how the 4CL genes evolve and functionally diverge and whether natural selection plays a role in their evolution have been poorly studied. In this study, we analyzed nucleotide divergence in the 4CL genes from 16 species and used likelihood methods with various evolutionary models to investigate potential patterns of positive selection.

## Methods

### Sequence data collection

All known and reported 4CL protein-coding sequences from dicots, monocots, and gymnosperms (loblolly pine) were retrieved from the National Center for Biotechnology Information (NCBI). In total, 42 4CL protein sequences from 16 species were collected and are listed in Table S1.

### Phylogenetic analysis

The 4CL protein-coding sequences were aligned using the program CLUSTALW implemented in MEGA5 (Tamura et al. 2011) and manually edited. Highly variable regions, indels, and gaps were excluded. A phylogenetic tree was constructed using MEGA5 with the neighbor-joining (NJ) method. The reliability of the branches was evaluated by 1000 bootstrap replicates.

### Test for selection

The nonsynonymous–synonymous substitution rate ratio ( $\omega = dN/dS$ ) provides a measure of the selective pressure at the protein level, where a  $\omega$  of 1, <1, or >1 indicates neutral evolution, purifying selection, or positive selection, respectively. The hypothesis of positive selection was tested using the CODEML program in the PAML v4.3b package (Yang 2007). Three approaches, branch, site, and branch-site models, incorporated into the program were used. In the lineage-specific selection analyses, we employed the recently developed dynamic programming procedure to search for the optimal branch-specific model that had a likelihood equal to or close to the global maximum likelihood for all of the possible models (Zhang et al. 2011). In the site-specific selection analyses, the dataset was fitted to three pairs of codon substitution models (M2a vs. M1a, M3 vs. M0, and M8 vs. M7). The branch-site model A was used to detect positively selected sites along the branches that showed elevated  $\omega$  ratios.

The sites under positive selection were identified by the Bayes Empirical Bayes (BEB) approach.

## Results and Discussion

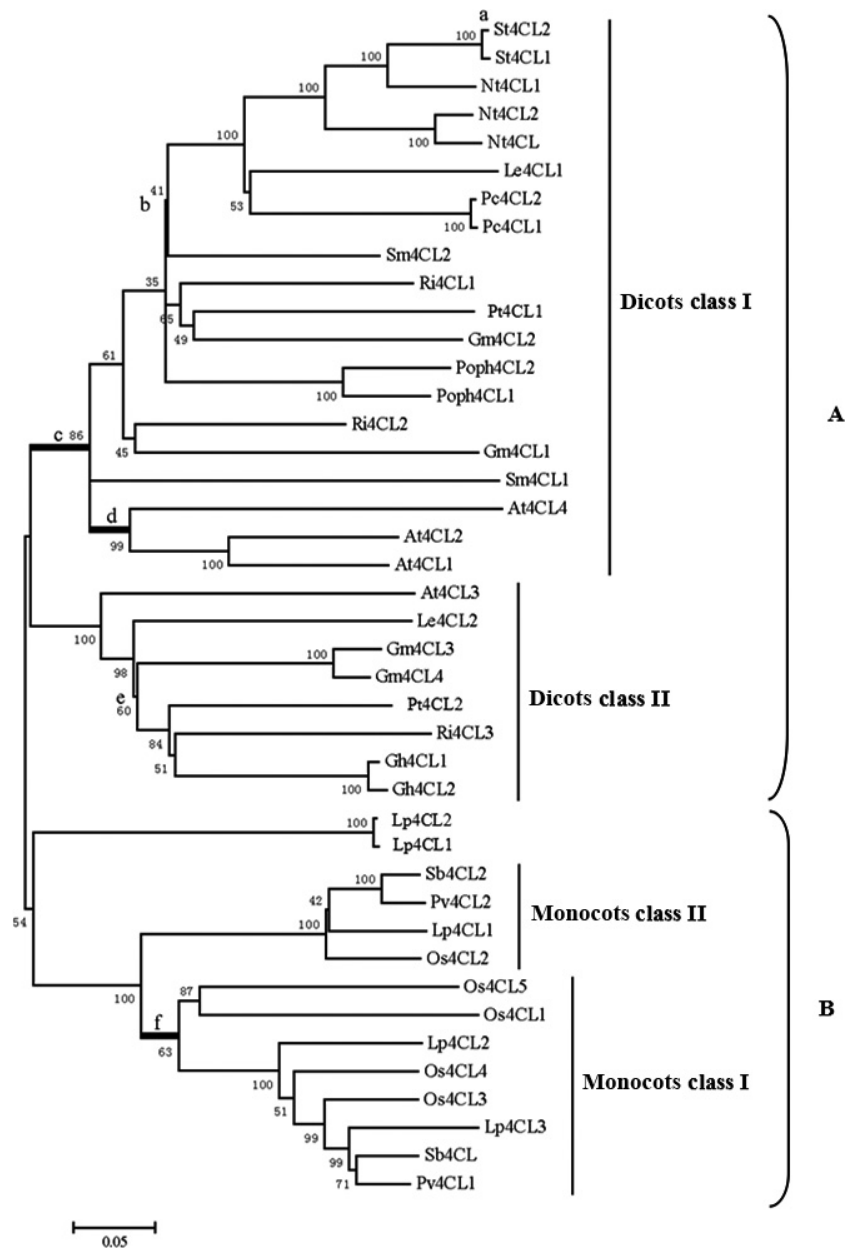
### Angiosperm 4CL gene phylogeny

The conserved protein-coding sequences of 42 4CLs from 16 species were used to reconstruct a phylogenetic tree. Analysis revealed that all of the 4CL genes fell into one of two general groups: A and B (Fig. 1). Group A contains representatives from all of the available dicots, including verified 4CL sequences from Arabidopsis, poplar and soybean. The monocot 4CL isoenzymes in group B form a highly supported monophyletic group and are thus separated from the dicot isoforms. The gymnosperm 4CLs, the loblolly pine isoforms Lp4CL1 and Lp4CL2, also formed a separate cluster that was closest to the monocot isoenzymes.

### The functional divergence of the 4CL gene family

The dicot 4CLs can be divided into two distinct groups that are designated dicots class I and dicots class II (Fig. 1). Previous studies have demonstrated that 4CL genes in dicots class I are associated with lignin accumulation, while dicots class II 4CLs are involved in the metabolism of other phenolic compounds, such as flavonoids. For example, the genes *Pt4CL1*, *At4CL1*, *At4CL2*, *At4CL4*, *Gm4CL1*, and *Gm4CL2* in dicots class I are involved in lignin formation (Hu et al. 1998; Ehlting et al. 1999; Lindermayr et al. 2002). However, the genes *Pt4CL2*, *At4CL3*, and *Gm4CL4* in dicots class II are believed to play a role in flavonoid biosynthesis (Uhlmann and Ebel 1993; Hu et al. 1998; Ehlting et al. 1999).

The 4CLs from monocots can also be classified into two groups, which are designated monocots class I and monocots class II (Fig. 1). The 4CL genes in monocots class I are associated with lignin accumulation. For example, *Pv4CL1* in monocots class I is the key 4CL isoenzyme involved in lignin biosynthesis because RNA interference of *Pv4CL1* reduces the activity of extractable 4CL by 80% leading to a reduction in lignin content and a decrease in the guaiacyl unit composition (Xu et al. 2011). The *Os4CL3* gene in the same group is also involved in lignin biosynthesis because suppression of *Os4CL3* expression results in significant lignin reduction, retarded growth and other morphological changes (Gui et al. 2011). However, the genes in monocots class II (Fig. 1) are likely to participate in the flavonoid biosynthetic pathway. For example, based on phylogenetic analysis, Xu et al. (2011) hypothesized that *Pv4CL2* in



**Figure 1.** Phylogenetic relationship between 4CL genes from angiosperms based on the neighbor-joining method. The branch lengths are proportional to distances, and the values at the interior nodes are the bootstrap percentages derived from 1000 replicates. The six branches potentially under positive selection are indicated as a, b, c, d, e, and f, respectively.

monocots class II mainly participates in the flavonoid biosynthesis pathway in switchgrass. Recent research (Sun et al. 2013) has demonstrated that the primary function of *Os4CL2* is to channel the activated 4-coumarate to chalcone synthase and subsequently to different branched pathways of flavonoid secondary metabolism leading to flower pigments and UV protective flavonols and anthocyanins. The remarkable functional diversity of not only dicot but also monocot 4CLs suggests that 4CL may be subject to positive Darwinian selection.

### Evolutionary patterns among lineages and among sites

To test the hypothesis that positive selection acts on 4CLs, we applied branch-specific models to the 4CL dataset. It was clear that 40RM (40 ratio model) with 40 different  $\omega$  ratios was the optimal branch model (Table S2). The six branches where  $\omega$  was  $>1$  were defined as branches a, b, c, d, e, and f, respectively (Fig. 1). To examine whether the  $\omega$  ratio for each branch was significantly greater than the background ratio, the log-likelihood values were calculated

from two-ratio models that assigned the ratios  $\omega_a, \omega_b, \omega_c, \omega_d, \omega_e,$  and  $\omega_f$  to branches  $a, b, c, d, e,$  and  $f,$  and the ratio  $\omega_0$  was assigned to all other branches. All of these two-ratio models were individually compared with the one-ratio model (M0). The one-ratio model, which assumes the same  $\omega$  parameter for the entire tree, yielded a log-likelihood value of -28951.48 with an estimated  $\omega_0$  of 0.089 (Table 1). The low average ratio indicated the dominating role of purifying selection in the evolution of the 4CL genes. The two-ratio models for branches  $a, c, d,$  and  $f$  fit the data significantly better than the one-ratio model

(Table 2), resulting in the rejection of the null hypothesis that the 4CL genes evolved at constant rates along the branches. To test whether the six  $\omega$  ratios were significantly higher than 1, we calculated the log-likelihood values using the two-ratio models with  $\omega_a, \omega_b, \omega_c, \omega_d, \omega_e,$  and  $\omega_f$  fixed to 1 (Table 1). The likelihood ratio tests were also implemented for comparing each two-ratio model and its corresponding fixed two-ratio model. The likelihood ratio tests in Table 2 revealed that the  $\omega$  ratios for branches  $a, b, c, d, e,$  and  $f$  were not significantly greater than one. We therefore conclude that the evolution of the

**Table 1.** Log-likelihood values and parameter estimates for the 4CL genes.

Model	$\rho$	lnL	Parameter estimates	Positively selected sites
M0: one ratio	1	-28951.48	$\omega_0 = 0.089$	None
Branch specific models				
Two ratios (branch a)	2	-28949.50	$\omega_0 = 0.089, \omega_a = 3.463$	
Two ratios (fixed $\omega_a = 1$ )	1	-28949.58	$\omega_0 = 0.089, \omega_a = 1$	
Two ratios (branch b)	2	-28950.10	$\omega_0 = 0.088, \omega_b = 0.273$	
Two ratios (fix $\omega_b = 1$ )	1	-28950.49	$\omega_0 = 0.088, \omega_b = 1$	
Two ratios (branch c)	2	-28948.26	$\omega_0 = 0.088, \omega_c = 2.270$	
Two ratios (fixed $\omega_c = 1$ )	1	-28948.29	$\omega_0 = 0.088, \omega_c = 1$	
Two ratios (branch d)	2	-28948.76	$\omega_0 = 0.088, \omega_d = 0.574$	
Two ratios (fixed $\omega_d = 1$ )	1	-28948.79	$\omega_0 = 0.088, \omega_d = 1$	
Two ratios (branch e)	2	-28950.04	$\omega_0 = 0.088, \omega_e = \infty$	
Two ratios (fixed $\omega_e = 1$ )	1	-28950.17	$\omega_0 = 0.088, \omega_e = 1$	
Two ratios (branch f)	2	-28939.38	$\omega_0 = 0.088, \omega_f = \infty$	
Two ratios (fixed $\omega_f = 1$ )	1	-28940.20	$\omega_0 = 0.088, \omega_f = 1$	
Sites-specific models				
M1: neutral (K = 2)	1	-28731.92	$\rho_0 = 0.946 (\rho_1 = 0.054)$	Not allowed
M2: selection (K = 3)	3	-28731.92	$\rho_0 = 0.946, \rho_1 = 0.011$ $(\rho_2 = 0.042), \omega_2 = 1$	None
M3: discrete (K = 2)	3	-28260.60	$\rho_0 = 0.520 (\rho_1 = 0.480)$ $\omega_0 = 0.020, \omega_1 = 0.175$	None
M3: discrete (K = 3)	5	-28136.31	$\rho_0 = 0.408, \rho_1 = 0.456, (\rho_2 = 0.136)$ $\omega_0 = 0.010, \omega_1 = 0.108, \omega_2 = 0.320$	None
M7: beta	2	-28118.00	$\rho = 0.542, q = 4.443$	Not allowed
M8: beta and $\omega$	4	-28116.60	$\rho_0 = 0.994, \rho = 0.570, q = 5.015$ $(\rho_1 = 0.006), \omega = 1$	None
Branch-site model A				
Model a	3	-28731.43	$\rho_0 = 0.912, \rho_1 = 0.052$ $(\rho_2 + \rho_3 = 0.036), \omega_2 = 9.870$	None
Model fixed $\omega_a$	2	-28731.55	$\rho_0 = 0.725, \rho_1 = 0.041$ $(\rho_2 + \rho_3 = 0.234), \omega_2 = 1$	Not allowed
Model c	3	-28719.14	$\rho_0 = 0.885, \rho_1 = 0.051$ $(\rho_2 + \rho_3 = 0.064), \omega_2 = 15.947$	82C 291S (at $P > 0.95$ ) 379M 423T (at $P > 0.99$ )
Model fixed $\omega_c$	2	-28724.93	$\rho_0 = 0.780, \rho_1 = 0.046$ $(\rho_2 + \rho_3 = 0.155), \omega_2 = 1$	Not allowed
Model d	3	-28723.32	$\rho_0 = 0.922, \rho_1 = 0.053$ $(\rho_2 + \rho_3 = 0.024), \omega_2 = 14.873$	79V 202S 211S (at $P > 0.95$ )
Model fixed $\omega_d$	2	-28726.61	$\rho_0 = 0.825, \rho_1 = 0.047$ $(\rho_2 + \rho_3 = 0.128), \omega_2 = 1$	Not allowed
Model f	3	-28716.01	$\rho_0 = 0.780, \rho_1 = 0.045$ $(\rho_2 + \rho_3 = 0.160), \omega_2 = \infty$	65L 69E 181I 223L 234K 239K (at $P > 0.95$ )
Model fixed $\omega_f$	2	-28720.97	$\rho_0 = 0.349, \rho_1 = 0.020$ $(\rho_2 + \rho_3 = 0.631), \omega_2 = 1$	Not allowed

4CL genes in angiosperms is dominated by purifying selection.

Because the branch model test averages the  $\omega$  ratios across all of the sites and is a very conservative test for positive selection, we applied site-specific models to the 4CL dataset. The log-likelihood values and the parameter estimates under models with variable  $\omega$  ratios among the sites are listed in Table 1. Two site classes (M3,  $K = 2$ ) fit the data significantly better than one site class (M0) by 690.88 log-likelihood units revealing significant variation in the selective pressure on the sites. However, none of the site-specific models allowed for the presence of positively selected sites, such as M2a (selection), M3 (discrete), and M8 (beta and  $\omega$ ), suggesting the existence of positively selected sites with  $\omega > 1$ . The majority of the sites in the 4CL sequences appear to be under strong selective constraints.

### Evidence for positive selection on lignin-related 4CL genes

Positive selection is difficult to detect because it often operates episodically on just a few amino acid sites and

**Table 2.** Likelihood ratio statistics ( $2\Delta\ln L$ ) for testing branch hypothesis.

	M0 (one ratio)	Fixed $\omega_a = 1$	Fixed $\omega_b = 1$	Fixed $\omega_c = 1$	Fixed $\omega_d = 1$	Fixed $\omega_e = 1$	Fixed $\omega_f = 1$
$\omega_a$ free	3.96*	0.16					
$\omega_b$ free	2.76		0.78				
$\omega_c$ free	6.44*			0.06			
$\omega_d$ free	5.44*				0.06		
$\omega_e$ free	2.88					0.26	
$\omega_f$ free	24.2**						1.64

\*Significant ( $P < 0.05$ ,  $\chi^2 = 3.84$ ).

\*\*Extremely significant ( $P < 0.01$ ,  $\chi^2 = 6.63$ ).

purifying selection may mask the signal. Branch-site models can detect positive selection that affected a small number of sites along prespecified lineages. We used branch-site model A to test the hypothesis. As detailed in Table 3, branch-site model A using branch *c* as the foreground branch (MAc) resulted in a significantly better fit to M1a ( $2\Delta\ln L = 25.56$ ,  $df = 2$ ,  $P < 0.00$ ) and to null model A for branch *c* ( $2\Delta\ln L = 11.58$ ,  $df = 1$ ,  $P < 0.00$ ) (Table 3). This result also suggested that 5.1% of amino acids are under positive selection in lineage *c* with  $\omega = 15.95$  (Table 1). Branch-site model A using branch *d* as the foreground branch (MA<sub>d</sub>) provided a significantly better fit to M1a ( $2\Delta\ln L = 17.2$ ,  $df = 2$ ,  $P < 0.00$ ) and the null model A for branch *d* ( $2\Delta\ln L = 6.58$ ,  $df = 1$ ,  $P < 0.00$ ) (Table 3). This result also suggested that 5.3% of the protein sites are under positive selection in lineage *d* with  $\omega = 14.873$  (Table 1). When the analysis was repeated with branch *f* as the foreground branch (MA<sub>f</sub>), model A was much more realistic and fit the data significantly better than M1a ( $2\Delta\ln L = 31.82$ ,  $df = 2$ ,  $P < 0.00$ ) and the null model A for branch *f* ( $2\Delta\ln L = 9.92$ ,  $df = 1$ ,  $P < 0.01$ ) (Table 3), which suggested that 4.5% of the amino acids are under positive selection in lineage *f* with  $\omega = \infty$  (Table 1). Model A using branch *a* as the foreground branch (MA<sub>a</sub>) did not fit the data better than the two null models in test 1 and test 2 (Table 3). These evidences are sufficient to support the positive selection hypothesis on lineages *c*, *d*, and *f*.

Based on the BEB method, four and six candidate sites for positive selection were identified in dicots and monocots, respectively (Table 1). These positively selected sites are labeled in Figure 2. Sites 181I, 202S, 211S, 223L, 234K, and 239K are located in the substrate-binding pocket, and 379M and 423T are located in the in catalytic centers. Sites 65L, 69E, 79V, and 82C are located between the conserved sequence motifs A2 and A3, which form a phosphate-binding loop. Site 291S is close to motif A6, which is important for the formation of a stable tertiary structure. Thus, amino acid substitutions in these positively selected sites in the 4CL genes might influence the 4CL substrate specificity, activity, or secondary structure, which would in turn have a profound effect on 4CL's function.

**Table 3.** Likelihood ratio statistics ( $2\Delta\ln L$ ) for testing branch-site hypothesis.

	M1a	Branch-site MAa ( $\omega_a = 1$ )	Branch-site MAc ( $\omega_c = 1$ )	Branch-site MA <sub>d</sub> ( $\omega_e = 1$ )	Branch-site MA <sub>f</sub> ( $\omega_e = 1$ )
MAa	0.98 (0.61)	0.24 (0.89)			
MAc	25.56 (2.82E-07)**		11.58 (6.67E-04)**		
MA <sub>d</sub>	17.2 (1.84E-04)**			6.58 (1.03E-02)*	
MA <sub>f</sub>	31.82 (1.23E-07)**				9.92 (7.01E-03)**

\*Significant ( $P < 0.05$ ).

\*\*Extremely significant ( $P < 0.01$ ).

**A2**  
**37** FRGATATAANPFFTPAEIAKQAKASNTKLIITEARYVDKIKVIVCIDDNVEISPDDEVVALPY

**Box I**  
**100** SSGTTGLPKGVMLTHKGLVTSVAQQVDGENPNLYFHSDDVILCVLPMFHIIYALNSIMLCG  
**A3**

**160** LRVGAAILMPKFEINLLELQRCKVTVAPMVPPIVLAIAKSETEKYDLSIRVVKSGAAP

**223** LGKELEDAVNAKFPNAKLGQGYGMTEAGVLAMSLGFAKEPFPVKSGACGTVVRNAEM

**Box II**  
**280** KIVDPDTGDSL SRNQPEICIRGHQIMKGYLNNPAATAETIDKDGWLHTGDIGLIDDELDF  
**A6**

**341** IVDRLKELIKYKGFQVAPAELEALLIGHPDITDVAVAMKEEAAGEVPVAFVVKSKDLS  
**A10**

**402** DVKQFVSKQVVFYKRINKVFFTESIPKAPSGKILRKDLRAKL

**Figure 2.** The deduced amino acid sequence for Arabidopsis At4CL1 referred to in this article. The residues involved in hydroxycinnamate binding are indicated with stars, while those involved in enzymatic functions are labeled with triangles (Hu et al. 2010). The bold-type letters indicate conserved motifs (Gulick 2009), while those on a gray background indicate positively selected sites.

We have demonstrated that 4CL genes in branches *c* and *f*, which are associated with lignin accumulation, are under positive selection. Interestingly, positive selection is also detected at the *At4CL* genes in branch *d*. However, the role of these proteins in lignin formation is similar to other proteins from dicots class I (Hu et al. 1998; Ehltling et al. 1999). We hypothesize that positive selection on the *At4CL* genes may be related to functional specialization.

### Selective constraints on flavonoid-related 4CLs in dicots

The 4CL genes involved in flavonoid biosynthesis (dicots class II and monocots class II, Fig. 1) have been largely conserved during plant evolution, suggesting that they are constrained by purifying selection. Land plants evolved from green algae in the mid-Ordovician over 450 million years ago (Langdale 2008). After arriving in terrestrial environments, the pioneering land plants were confronted with several major challenges such as ultraviolet irradiation, desiccation stress. The presence of flavonoid in the earliest land plants and the associated ability to resist UV irradiations made survival on land possible for the plants (Rozemaa et al. 2002). Flavonoid evolved prior to the lignin pathway. For example, bryophytes do not synthesize lignin, but accumulate soluble phenylpropanoids, such as flavonoids and lignans (Weng and Chapple 2010). Flavonoids accumulate in the epidermal layer of extant plants, which has been shown to absorb over 90% of UV-B radiation (Stafford 1991). These evidences suggested that the ancestral role of 4CL was to participate in the flavonoid

biosynthesis and that this role was maintained in the adaption to a terrestrial environment.

### Conclusions

4CLs play important roles in both lignin and flavonoid biosynthesis. 4CLs that play a role in lignin biosynthesis are subject to positive selection. This positive selection resulted in a functional divergence after the monocot–dicot split approximately 200 million years ago. Positive selection could have been involved in the early stages of the evolution of the 4CL genes; 4CL rapidly evolves after speciation events. Strong purifying selection operates on the novel 4CL genes to maintain the protein's existing function. Based on the BEB method, four and six candidate sites for positive selection were identified in dicots and monocots, respectively (Table 1). Most of the positively selected sites are located in the substrate-binding pocket and the catalytic centers (Fig. 2). Therefore, amino acid replacements in these sites might imply a neofunctionalization. The result is in agreement with our findings that 4CL genes functionally diversified in angiosperms (Hu et al. 1998; Ehltling et al. 1999; Gui et al. 2011; Xu et al. 2011; Sun et al. 2013). Although several positively selected sites were detected using the branch-site model, we find that the 4CL gene family as a whole experiences purifying evolution rather than pervasive selection throughout evolution. The 4CLs involved in flavonoid biosynthesis have been largely conserved during plant evolution and maintain the ancestral role in response to biotic or abiotic factors. These findings provide deeper insights into understanding the evolutionary



mechanisms of 4CL isoforms and their functional diversification.

## Acknowledgments

We thank Dr. Peng Chen and Dr. Liqiang Wang for reading and discussing the manuscript and Dr. Chengjun Zhang for assistance with statistical analysis. This work was supported in part by grants from the National Transgenic Project (2009ZX08009-119B), the 111 Project (B08032), the 973 Specific Pre-project (2010CB134401), and the Youth Foundation of Jiangsu Province (BK20140417).

## Conflict of Interest

None declared.

## References

- Agati, G., C. Brunetti, M. Di Ferdinando, F. Ferrini, S. Pollastri, and M. Tattini. 2013. Functional roles of flavonoids in photoprotection: new evidence, lessons from the past. *Plant Physiol. Biochem.* 72:35–45.
- Ehrling, J. E., D. Buttner, Q. Wang, C. J. Douglas, I. E. Somssich, and E. Kombrink. 1999. Three 4-coumarate:coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. *Plant J.* 19:9–20.
- Gui, J., J. Shen, and L. Li. 2011. Functional characterization of evolutionarily divergent 4-coumarate:coenzyme A ligases in rice. *Plant Physiol.* 157:574–586.
- Gulick, A. M. 2009. Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. *ACS Chem. Biol.* 4:811–827.
- Hamberger, B., and K. Hahlbrock. 2004. The 4-coumarate:CoA ligase gene family in *Arabidopsis thaliana* comprises one rare, sinapate-activating and three commonly occurring isoenzymes. *Proc. Natl Acad. Sci. USA* 101:2209–2214.
- Harding, S. A., J. Leshkevich, V. L. Chiang, and C. J. Tsai. 2002. Differential substrate inhibition couples kinetically distinct 4-coumarate:coenzyme A ligases with spatially distinct metabolic roles in quaking aspen. *Plant Physiol.* 128:428–438.
- Hu, W. J., A. Kawaoka, C. J. Tsai, J. Lung, K. Osakabe, H. Ebinuma, et al. 1998. Compartmentalized expression of two structurally and functionally distinct 4-coumarate:CoA ligase genes in aspen. *Proc. Natl Acad. Sci. USA* 95:5407–5412.
- Hu, Y., Y. Gai, L. Yin, X. Wang, C. Feng, L. Feng, et al. 2010. Crystal structures of a *Populus tomentosa* 4-coumarate:CoA ligase shed light on its enzymatic mechanisms. *Plant Cell* 22:3093–3104.
- Langdale, J. A. 2008. Evolution of developmental mechanisms in plants. *Curr. Opin. Genet. Dev.* 18:368–373.
- Lee, D., K. Meyer, C. Clint, and C. Douglas. 1997. Antisense suppression of 4-Coumarate:coenzyme A ligase activity in *Arabidopsis* leads to altered lignin subunit composition. *Plant Cell* 9:1985–1998.
- Lindermayr, C., B. Mollers, J. Fliegmann, A. Uhlmann, and F. Lottspeich. 2002. Divergent members of a soybean (*Glycine max* L.) 4-coumarate:coenzyme A ligase gene family. *Eur. J. Biochem.* 269:1304–1315.
- Nakashima, J., F. Chen, L. Jackson, G. Shadle, and R. A. Dixon. 2008. Multi-site genetic modification of monolignol biosynthesis in alfalfa (*Medicago sativa*): effects on lignin composition in specific cell types. *New Phytol.* 179:738–750.
- Rozemaa, J., L. O. Bjorn, J. F. Bornman, A. Gaberscik, D. P. Hader, T. Trost, et al. 2002. The role of UV-B radiation in aquatic and terrestrial ecosystems—an experimental and functional analysis of the evolution of UV-absorbing compounds. *J. Photochem. Photobiol., B* 66:2–12.
- Saballos, A., S. E. Sattler, E. Sanchez, T. P. Foster, Z. Xin, C. Kang, et al. 2012. Brown midrib2 (*Bmr2*) encodes the major 4-coumarate:coenzyme A ligase involved in lignin biosynthesis in sorghum (*Sorghum bicolor* (L.) Moench). *Plant J.* 70:818–830.
- Souza, C. A., B. Barbazuk, S. G. Ralph, G. Bohlmann, B. Hamberger, and C. J. Douglas. 2008. Genome-wide analysis of a land plant-specific acyl:coenzymeA synthetase(ACS) gene family in *Arabidopsis*, poplar, rice and *Physcomitrella*. *New Phytol.* 179:987–1003.
- Stafford, H. A. 1991. Flavonoid evolution: an enzymic approach. *Plant Physiol.* 96:680–685.
- Sun, H., Y. Li, S. Feng, W. Zou, K. Guo, C. Fan, et al. 2013. Analysis of five rice 4-coumarate:coenzyme A ligase enzyme activity and stress response for potential roles in lignin and flavonoid biosynthesis in rice. *Biochem. Biophys. Res. Commun.* 430:1151–1156.
- Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28:2731–2739.
- Uhlmann, A., and J. Ebel. 1993. Molecular cloning and expression of 4-coumarate:coenzyme A ligase, an enzyme involved in the resistance response of soybean (*Glycine max* 1.) against pathogen attack. *Plant Physiol.* 102:1147–1156.
- Weng, J. K., and C. Chapple. 2010. The origin and evolution of lignin biosynthesis. *New Phytol.* 187: 273–285.
- Xu, B., L. L. Escamilla-Trevino, N. Sathitsuksanoh, Z. Shen, H. Shen, Y. H. Zhang, et al. 2011. Silencing of 4-coumarate:coenzyme A ligase in switchgrass leads to reduced

- lignin content and improved fermentable sugar yields for biofuel production. *New Phytol.* 192:611–625.
- Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24:1586–1591.
- Zhang, C., J. Wang, W. Xie, G. Zhou, M. Long, and Q. Zhang. 2011. Dynamic programming procedure for searching optimal models to estimate substitution rates based on the maximum-likelihood method. *Proc. Natl Acad. Sci. USA* 108:7860–7865.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Table S1.** Members of plant 4-coumarate:coenzyme A ligase (*4CL*) genes.

**Table S2.** Likelihood Ratio Statistics ( $2\Delta\ln L$ ) for Testing Method1 result.