

# The 4-coumarate:CoA ligase gene family in *Arabidopsis thaliana* comprises one rare, sinapate-activating and three commonly occurring isoenzymes

Björn Hamberger\* and Klaus Hahlbrock

Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, 50829 Cologne, Germany

Contributed by Klaus Hahlbrock, December 19, 2003

4-Coumarate:CoA ligase (4CL; EC 6.2.1.12) has a pivotal role in the biosynthesis of plant secondary compounds at the divergence point from general phenylpropanoid metabolism to several major branch pathways. In *Arabidopsis thaliana*, we have identified a previously undetected, fourth and final member of the *At4CL* gene family. The encoded enzyme, *At4CL4*, exhibits the rare property of efficiently activating sinapate, besides the usual 4CL substrates (4-coumarate, caffeate, and ferulate), indicating a distinct metabolic function. Phylogenetic analysis suggests an early evolutionary and functional divergence of three of the four gene family members, *At4CL2–4*, whereas *At4CL1* appears to have originated much later by duplication of its structurally and functionally closest relative, *At4CL2*. Various characteristics shared by all known plant 4CL genes, as well as by the encoded proteins, define and delimit the *At4CL* gene family and distinguish it from the closely related family of “*At4CL-like*” genes.

The 4-coumarate:CoA ligase family (4CL; EC 6.2.1.12) catalyzes the activation of 4-coumarate and a few related substrates to the respective CoA esters and thus channels the common, phenylalanine-derived building block into the otherwise widely distinct branches of general phenylpropanoid metabolism. These phenylpropanoid branch pathways generate various classes of natural (secondary) compounds with essential functions in plant development and environmental interactions, such as lignin for structural support, flavones and flavonols for UV protection, anthocyanins, chalcones and aurones as pigments for the attraction of pollinators and seed distributors, and isoflavonoids and furanocoumarins as phytoalexins for pathogen defense.

As far as analyzed, 4CL has been shown to occur in the form of multiple isoenzymes, which frequently (1–5), but not always (6, 7), exhibit distinct substrate affinities that appear to coincide with specific metabolic functions. Of the four naturally occurring, potential substrates illustrated in Fig. 1, all except sinapate are efficiently activated by numerous 4CLs from various sources, whereas sinapate has so far been converted with appreciable efficiency, apart from crude or partially purified protein extracts (2, 5, 8–10), by just one pure heterologously expressed isoenzyme, *Gm4CL1* from soybean (*Glycine max* L.) (11). A recent combination of structural, biochemical, and mutational analyses using *At4CL2* from *Arabidopsis thaliana* as a test case revealed that the substrate-binding pocket of this and all other 4CLs with the same specificity-determining amino acid code is too small to accommodate sinapate (12, 13). Directed mutations enlarging this pocket to a size similar to that realized in *Gm4CL1* indeed led to the expected broadening of the substrate specificity, including the conversion of sinapate (13). *In silico* studies in the same context identified one additional, functionally undefined, “4CL-like” (CLL) protein in *A. thaliana* that contained a similar large substrate-binding pocket and was thereupon shown to also be a true sinapate-activating 4CL (13).

This latter discovery, in the context of our long-term interest in understanding the functional diversity of phenylpropanoid metabolism in conjunction with its genetic background in *A.*

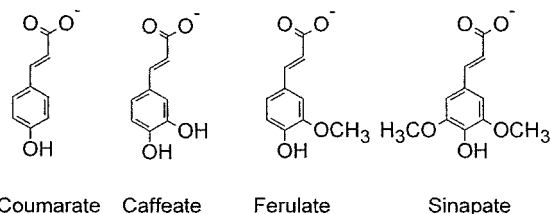


Fig. 1. Chemical structures of the four naturally occurring 4CL substrates investigated in this study.

*thaliana* (14), prompted us to further characterize this rare sinapate-activating *At4CL* and define its position within the *At4CL* gene family. Here, we verify its identity as a previously undetected fourth isoenzyme (*At4CL4*) in *A. thaliana*, compare its structural and putative functional characteristics with those of the other three family members, propose a sequence for the stepwise evolution of the four *At4CLs* as a basis for the diversification of phenylpropanoid metabolism, and discuss the results in relation to the substructure of the overall plant 4CL superfamily.

## Methods

**Bacterial Strains, Plasmids, Primer Sequences, and DNA Manipulation Techniques.** Standard DNA techniques were performed as described (15). For plasmid amplification and maintenance, the *Escherichia coli* strain DH5 $\alpha$  (GIBCO/BRL Life Technologies) was used. Protein expression was carried out by using the *E. coli* expression strain BL21(DE3) (Novagen).

cDNA from elicitor-treated suspension-cultured *A. thaliana* cells (16) was used as template for PCR amplification of full-length *At4CL4* cDNA. The *At4CL4* coding sequence, including the putative start codon, was amplified by using the oligonucleotide primers *At4CL4A* (CCACACATATGGT-GCTCCAACAACAAACGC) and *At4CL4B* (TCTTGCTC-GAGTTTAGAGCACATGGTTTCCAA). Two new restriction sites, for *NdeI* and *XhoI*, were introduced for cloning into the pET-30 expression plasmid (Novagen) and eliminating the endogenous stop codon. DNA sequences were determined on Applied Biosystems (Weiterstadt, Germany) Prism 377 and 3700 sequencers by using BigDye-terminator chemistry. Premixed reagents were from Applied Biosystems. Oligonucleotides were purchased from Qiagen (Operon, Cologne, Germany).

**Purification of *At4CL4* and Enzyme Assay.** Expression and purification of the recombinant *At4CL4* protein carrying a C-terminal His<sub>6</sub> tag were performed as described in refs. 17 and 18 with

Abbreviations: 4CL, 4-coumarate:CoA ligase; CLL, 4CL-like.

\*To whom correspondence should be sent at the present address: Department of Botany, University of British Columbia, Vancouver, BC, Canada V6T 1Z4. E-mail: bjoernh@interchange.ubc.ca.

© 2004 by The National Academy of Sciences of the USA

minor modifications. Purified proteins were separated by polyacrylamide gel electrophoresis in the presence of 0.1% SDS (SDS/PAGE), transferred onto nylon membranes, and detected by using an antiserum raised against the parsley (*Petroselinum crispum*) Pc4CL (19).

**Comparison and Analysis of Protein and Nucleotide Sequences.** The following 4CL sequences were used for phylogenetic analyses (GenBank accession numbers in parentheses): *A. thaliana* At4CL4 (AAM19949), *Capsicum annuum* Ca4CL (AAG43823), *G. max* Gm4CL1 (AAL98709), Gm4CL2 (AAC97600), Gm4CL3 (AAC97599), Gm4CL4 (CAC36095), *Lolium perenne* Lp4CL1 (AAF37732), Lp4CL2 (AAF37733), Lp4CL3 (AAF37734), *Populus balsamifera* subsp. *trichocarpa* × *Populus deltoides* Pb4CL3 (AAK58908), Pb4CL4 (AAK58909), *Populus tomentosa* PtPo4CL (AAL56850), *Rubus idaeus* Ri4CL1 (AAF91310), Ri4CL2 (AAF91309), Ri4CL3 (AAF91308), and *Streptomyces coelicolor* Sc4CL (CAB95894). All other 4CL sequences were the same as compiled elsewhere (18).

Sequence comparison of proteins was performed by using the GAP program of the Genetics Computer Group (GCG, default parameters). cDNA sequences were compared by using the program DIALIGN (Genomatix Software, Munich).

Sequence alignments, followed by manual optimization, were achieved by using the PILEUP program (GCG) package, version 10.0, the BLOSUM62 matrix, and default parameters (20). On the basis of these alignments, a maximum parsimony analysis (21) was performed by using the PAUPSEARCH program (GCG; PAUP version 4.01 embedded; Smithsonian Institution, Washington DC). The most parsimonious tree was found by a heuristic search using the branch swapping-tree bisection and reconnection (TBR) algorithm (22) and 500 bootstrap replications (23). The topology of the tree was confirmed by using an exhaustive branch-and-bound (BB) algorithm and 500 bootstrap replications on a small subset of 12 representative sequences.

The PLACE Database (24) was used to screen the At4CL promoter regions for perfect W-box elements (TTGACC; ref. 25).

## Results

**Identification, Heterologous Expression, and Biochemical Characterization of At4CL4.** Analysis of the publicly accessible *A. thaliana* genome sequence data (National Center for Biotechnology Information) revealed the presence of one further At4CL gene (At4CL4) in addition to the three previously reported isoforms, At4CL1–3 (18, 26). In contrast to At4CL1–3, no expressed sequence tag (EST) clones existed for At4CL4, indicating comparatively low expression rates under the conditions used. We therefore cloned the At4CL4 cDNA from a library derived from elicitor-treated suspension-cultured *A. thaliana* cells. The encoded enzyme was heterologously expressed as a C-terminally His<sub>6</sub>-tagged protein in *E. coli*, and its immunogenic identity as 4CL was demonstrated by using a previously generated 4CL-specific antiserum (19).

The enzyme-kinetic properties of the purified protein (Table 1) verified its biochemical function as a *bona fide* 4CL, notably with the rare peculiarity that At4CL4, in contrast to all but one (11) of numerous previously characterized 4CLs from a wide range of plant species, efficiently converted sinapate to the corresponding CoA ester. In fact, according to the  $K_m$  and the  $V_{max}/K_m$  values, sinapate, together with its close structural relative ferulate, proved to be the most efficient of the four tested naturally occurring 4CL substrates (Fig. 1). Three selected, less polar, yet sterically similarly demanding (3,4-, 3,5-, and 3,4,5-methoxylated) cinnamate derivatives, all of which lacked the characteristic 4-hydroxy group of the four accepted substrates, were very poorly converted at negligible rates, and the

**Table 1. Kinetic properties of At4CL4 in vitro**

Substrate	$K_m$ , $\mu\text{M}$	Specific activity, $V_{max}/\text{mg}$	$V_{max}/K_m$
4-Coumarate	432	100	0.3
Caffeate	186	187	1.1
Ferulate	26	153	6.6
Sinapate	20	105	6.7

All data are mean values from two independent experiments with separately expressed proteins. See Fig. 1 for chemical structures of substrates.  $V_{max}$  is in nkat (1 kat = 1 mol·s<sup>-1</sup>).

unsubstituted parent compound, cinnamate, was not converted at all within experimental error (data not shown).

**Comparative Analysis and Definition of the Complete At4CL Gene Family.** Table 2 shows the nucleotide and amino acid sequence similarities both within the At4CL gene family and between each of the four At4CL isoforms and their next closest relative (At1g20510), one of several functionally undefined AtCLL genes (18, 26). According to these data, At4CL4 is related to each of the class I members (18) at a similar evolutionary distance, whereas At4CL1 and At4CL2 are much more closely related to one another and At4CL3 is most distantly related to all other isoforms. Importantly, there is a clear-cut distinction between the At4CL gene family and even the most closely related AtCLL gene sequence, At1g20510.

These differences in sequence relationship among the four At4CL genes are fully matched at the level of intron/exon structure (Fig. 2). Again, At4CL1 and At4CL2 are very closely related, At4CL4 has an intermediate position, and At4CL3 is a structurally more distant family member. Furthermore, all four At4CL genes contain, in the TATA-proximal region, at least one copy each of the two promoter elements, boxes P and L (27, 28), that contain a putative MYB-binding site. These elements are characteristic without exception of all genes involved in general phenylpropanoid metabolism investigated so far and have been described for a wide range of plant species (see below). The less frequently occurring box A (28) is not present in any one of the four At4CL gene promoters without at least one (At4CL3) or two (At4CL4) mismatches. An absolute singularity among all known 4CL genes is the occurrence of three perfect TATA-proximal W-box sequences (25) in the At4CL4 gene promoter, two of which have recently been shown to fulfil the operational W-box criterion of WRKY transcription factor binding (14).

At the protein level, a recent comparative analysis of the 4CL substrate specificity-determining amino acids, based on 3D homology modeling, revealed the presence of 12 functionally essential amino acid residues proposed to form the substrate-binding pocket of At4CL2 (13). In the same study, amino acid comparison of all four isoenzymes, including At4CL4 [tentatively designated as CLL (At3g21230)] and Gm4CL1 showed a matching amino acid signature for all 4CLs with a characteristic

**Table 2. Comparison of nucleotide and (in parentheses) amino acid sequence identities (%) among the four At4CL isoforms and their closest AtCLL relative in the *A. thaliana* genome**

	At4CL1	At4CL2	At4CL3	At4CL4
At4CL1	—	74 (83)	54 (61)	63 (66)
At4CL2	—	—	54 (62)	62 (65)
At4CL3	—	—	—	52 (58)
At1g20510	40 (41)	37 (38)	34 (40)	39 (38)

cDNA coding regions were used for nucleotide sequence comparison.

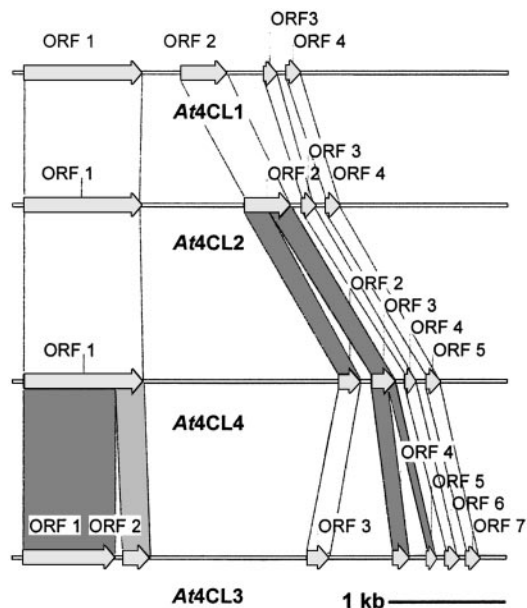


Fig. 2. Comparison of exon/intron structures of *At4CL1-4*.

sinapate conversion-specifying deletion in both *At4CL4* and *Gm4CL1*.

Neither the amino acid signature of the substrate-binding pocket (as well as the 4CL-characteristic highly conserved peptide motif, box I, and the presence of box II, which flank the substrate-specificity code) nor the equally characteristic promoter boxes P and L were found in the next closest relative, *At1g20510* (see above). Together, these data demonstrate (i) that *At4CL4* is a true member of the *At4CL1-4* gene family and (ii) that the family is confined to these four isoforms.

**Genomic Distribution of the *At4CL* Gene Family.** The particularly high sequence similarity between *At4CL1* and *At4CL2* suggests a much more recent gene duplication event than deemed likely for all other possible *At4CL* pairings. Fig. 3 indicates that such a recent duplication event would most probably have involved a sizeable genome segment (29). One copy each of this segment is located on chromosomes I and III and contains, in addition to *At4CL1/At4CL2* and several ORFs of unknown function, at least four functionally identified genes in identical relative positions and orientation. All of the encoded proteins share the same high sequence similarity (>80% identity) as shown for *At4CL1* and *At4CL2* (Table 2).

**Phylogenetic Analysis of the Plant 4CL Superfamily.** All available amino acid sequences of the hitherto established plant 4CLs were used for phylogenetic analysis of the 4CL superfamily. Fig.

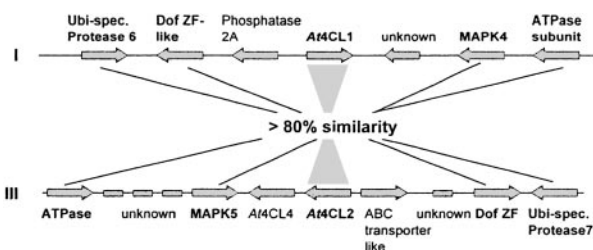


Fig. 3. Relative positions of sequence-related genes on inversely duplicated segments of *A. thaliana* chromosomes I and III. Bold letters indicate crosswise functional as well as close sequence relationship. Ubi-spec., ubiquitin-specific.

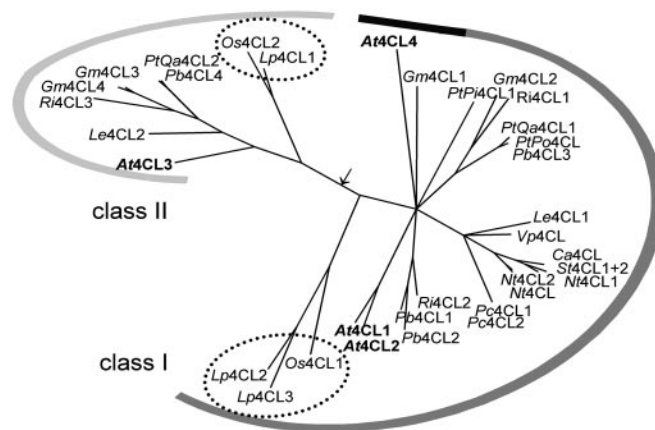


Fig. 4. Unrooted phylogenetic tree of all presently known plant 4CL isoenzymes. The putative root position is marked by an arrow. Contiguous graduated lines indicate class membership (dark gray = sinapate-activating isoenzymes). *At*, *Arabidopsis thaliana*; *Ca*, *Capsicum annuum*; *Gm*, *Glycine max*; *Le*, *Lithospermum erythrorhizon*; *Lp*, *Lolium perenne*; *Nt*, *Nicotiana tabacum*; *Os*, *Oryza sativa*; *Pb*, *Populus balsamifera* subsp. *trichocarpa* × *Populus deltoides*; *Pc*, *Petroselinum crispum*; *PtPi*, *Pinus taeda*; *PtPo*, *Populus tomentosa*; *PtQa*, *Populus tremuloides*; *Ri*, *Rubus idaeus*; *St*, *Solanum tuberosum*; *Vp*, *Vanilla planiflora*. Not shown are "*Pc4CL3*" and "*Nt4CL3*" (see text), both of which were deduced from partial nucleotide sequences and thus insufficiently characterized. Dotted circles enclose monocotyledonous representatives.

4 shows the most parsimonious unrooted tree generated, which clearly distinguishes between the previously defined class I and class II clades (18) and further supports the notion that *At4CL1* and *At4CL2* are evolutionarily closely linked, whereas *At4CL1/2*, *At4CL3*, and *At4CL4* are more distant relatives. Among the four, *At4CL3* is the only paralogue representing the comparatively infrequently occurring class II. All *At4CL* proteins formed a third, much more distantly related, clade at a considerable distance from the depicted tree (data not shown), confirming and extending previous results (13, 26) and further substantiating the confinement of the *At4CL* family to isoforms 1-4.

Particularly noteworthy with regard to the unusual substrate specificity of *At4CL4* is its close spacing in the tree with the only other known sinapate-activating ortholog, *Gm4CL1*, and the fairly isolated position of these two outsiders within class I. Similarly remarkable is the clear-cut separation of all known monocotyledonous (dotted circles in Fig. 4) from dicotyledonous enzymes in each class.

## Discussion

The identification and biochemical characterization of *At4CL4* as an exceptional type of 4CL isoenzyme with unusual substrate preference *in vitro* has considerably broadened the putative functional spectrum of the *At4CL* family *in vivo*. In this regard, *At4CL* is an extreme example of the majority of presently known 4CL families, including *Gm4CL*, *Le4CL*, *Os4CL*, *Pb4CL*, *Pt4CL*, and *Ri4CL* (Fig. 4), all of which exhibit a large structural, enzyme-kinetic, and apparent evolutionary diversity. The very opposite is exemplified by the much smaller and narrowly confined *St4CL*, *Pc4CL*, and *Nt4CL* families, which appear to consist of no more than two (*St4CL*) or three (*Pc4CL*, *Nt4CL*) nearly sequence-identical isoenzymes with indistinguishable substrate specificities (6, 7, 30). However, in all of these cases, with the single exception of *At4CL*, the true family sizes remain unknown until fully sequenced genomes are available.

*At4CL* now represents a definitely complete 4CL family, which can be fully described with regard to overall composition, delimitation, structural characteristics, evolutionary diversifica-



tion, and putative functional diversity. This task is greatly facilitated by the fact that *At4CL* belongs to the former group of 4CL families and consists of isoenzymes with widely differing degrees of structural and putative functional relationships to one another, including such distinct levels as substrate affinity *in vitro* (Table 1; ref 18), nucleotide and amino acid sequence (Table 2, Fig. 4), substrate-pocket design (13), promoter element composition, and, although not yet extensively explored, expression profile (*At4CL1–3*; ref. 18). The schematic depiction of the *At4CL* family (Fig. 4) was deliberately confined to a direct comparison with all other known plant 4CLs and hence does not include the next-closest relatives by sequence similarity, the functionally uncharacterized *AtCLL* proteins (18). All of these proteins form a separate branch at a considerable distance outside the class I and II 4CL branches (26, 31), thus clearly delimiting the *At4CL* family. However, one important modification of earlier tree representations (18, 26) is the reassignment of the former “*AtCLL*” protein, At3g21230, to the *At4CL* family and the consequential extension of this family from the previously reported three to four members.

The subdivision of most 4CL families into two classes, I and II, reflects not only the apparent evolutionary distances among the respective isoenzymes, particularly for mono- and dicotyledonous plants (Fig. 4), but also distinct metabolic functions (18). As far as functionally assigned, class I isoenzymes have been directly or indirectly associated with the biosynthesis of lignin and structurally related soluble or cell wall-bound phenylpropanoid derivatives, whereas class II isoenzymes have been associated with flavonoid biosynthesis. Accordingly, on the basis of expression studies and enzymatic properties, the three previously identified isoenzymes, *At4CL1/2* (class I) and *At4CL3* (class II), have been assigned to the lignin and flavonoid branches of phenylpropanoid metabolism, respectively (18). However, definitive and more refined functional assignments will now require comparative studies of all four *At4CL* family members, including gene-specific knock-out mutations and other specifically targeted approaches in combination with detailed morphological, histochemical, physiological, and biochemical phenotyping.

Considering the complexity and the extent of overlapping mRNA expression patterns that have been observed in various species for nearly all tested 4CL isoenzyme combinations (4, 7, 32, 33), including *At4CL1–3* (18), the pivotal position of 4CL in plant phenylpropanoid metabolism is probably reflected by a highly diversified metabolic grid (34–36). Moreover, the combination of distinct substrate preferences with differences in gene-promoter and protein fine structure may facilitate not only a large diversity of genetically programmed metabolic functions of the individual isoenzymes, including pathway-specific associations with biosynthetically related enzymes in metabolite channeling (37–39), but also regular or potential functional overlaps or mutual replacements upon malfunction or down-regulation, as has been investigated so far only for the class I isoforms (4, 40, 41).

The apparent specifically targeted evolution of *At4CL4* with its unique substrate preference suggests a special metabolic function, most probably related to sinapate activation, that could not be taken by any one of the other three isoenzymes. As a member of class I, *At4CL4* might be regarded as a lignin-related CoA ligase that channels highly substituted cinnamate derivatives into the biosynthesis of syringyl (S)-type lignin. However, this would contradict the presently favored model proposing linear biosynthetic routes to both guaiacyl (G)-type and S-type lignin (36), where 4CL acts at a relatively early stage before methylation. This model is supported by data indicating that *A. thaliana* 4-coumarate 3-hydroxylase (C3H) preferentially hydroxylates shikimate and quinate esters of coumarate, and that aspen caffeate *O*-methyltransferase (COMT) and *A. thaliana*

ferulate 5-hydroxylase (F5H) act on aldehydes and alcohols rather than on free acids (42–44). These observations would imply that 4CL in *A. thaliana* is not involved in ferulate, 5-hydroxyferulate, and sinapate activation en route to monolignol precursors, although radiolabeling studies indicated that ferulate and sinapate are incorporated into lignin in poplar (45). Thus, either alternatively or in addition, *At4CL4* may well have its major role elsewhere in phenylpropanoid metabolism, e.g., in the biosynthesis of soluble sinapate-containing phenolics. In any case, the metabolic complexity is likely to be greater than hitherto presumed, at least at the level of highly substituted cinnamate derivatives.

Whether similar sinapate-activating isoenzymes have evolved only in a few species or are too poorly expressed to be easily detected remains to be seen. One prediction is that plants lacking S-type lignin, such as conifers (e.g., pine), and plants with high proportions of 4-hydroxyphenyl (H)-type lignin, such as monocotyledonous plants (e.g., rice), either do not possess such an isoenzyme or require sinapate activation outside lignin biosynthesis. The ongoing rapidly progressing genome- and EST-sequencing projects (46–49) may soon provide the necessary basis for testing these alternatives.

The structural diversity of the *At4CL* family appears to be greatest, apart from introns, at the gene promoter level. The only common feature of all four gene promoters is the occurrence of the two cis-regulatory elements, boxes P and L, that have previously been shown to be a characteristic of plant genes related to general phenylpropanoid metabolism (27, 28, 50). These two boxes, together with a frequently, but not invariably, co-occurring third one, box A, were initially identified by “*in vivo* footprinting” as UV- and elicitor-responsive regions on the parsley *phenylalanine ammonia-lyase 1* (*PcPAL1*) gene promoter (27) and were assumed to be involved in the high degree of coordination that has always been observed for the expression of phenylpropanoid-biosynthetic genes in a large variety of conditions (28, 51–53). In fact, several enzymes of this pathway (54), including 4CL (55), were initially discovered on the basis of this extraordinary property. Here we show that the *At4CL4* gene promoter also contains both the P and L boxes at TATA-proximal positions (between –220 and –240 from the putative transcription start site, as well as at an atypical distance of –734) and thus fulfils this essential criterion of a gene from general phenylpropanoid metabolism. Remarkably, these two boxes occur in the same reverse order, relative to the model gene, *PcPAL1*, as realized on the *At4CL1–3* (18) and the initially studied *Pc4CL1–2* (52) gene promoters. Otherwise, however, the four *At4CL* gene promoters, in sharp contrast to the downstream coding regions, share no obvious sequence similarity with one another.

Although the functional significance of boxes P and L has not been tested in the particular case of *At4CL4*, their involvement in the coordinated induction, by endogenous as well as exogenous stimuli, of *At4CL4* with metabolically related enzymes is, by inference from previous results (18), highly probable. Less obvious is the role of the three W boxes on the *At4CL4* gene promoter, particularly in view of their possible functional overlap with the P and L boxes as elicitor (or pathogen) response elements (25). However, as each of the two types of element, boxes P/L as well as W, responds to several different kinds of stimulus, particularly multifaceted expression patterns might be expected for *At4CL4*. Whatever the significance of the unexpected W-box occurrence on the *At4CL4* gene promoter, their unique presence on this and not on any other known 4CL gene adds an interesting putative regulatory singularity to the exceptional biochemical properties of *At4CL4*.

The combined results on gene structure and genomic distribution of *At4CL1–4* (Figs. 2–4) suggest a sequence of three gene duplications and one functional loss during evolution of the

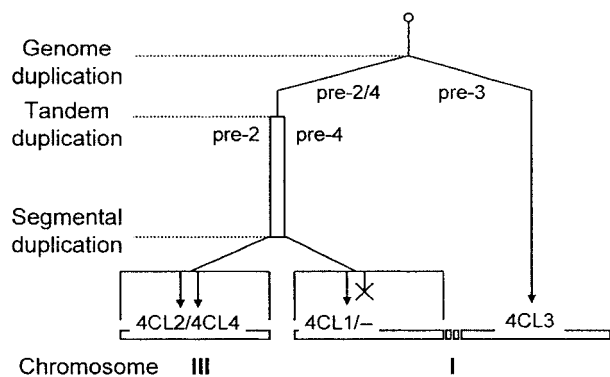


Fig. 5. Proposed scheme for the evolutionary divergence of *At4CL1-4*.

family (Fig. 5). According to this model, an early genome duplication event generated two copies of an ancestral gene, one of which evolved into *At4CL3* on chromosome I, whereas the other copy, on chromosome III, first underwent an early tandem duplication giving rise to the *At4CL4/At4CL2* precursors, and then a relatively recent segment duplication (29) back onto chromosome I, followed by loss of the *At4CL4* paralogue and fixation of *At4CL1* on this latter segment. The inferred, irregular timing of such a sequence of events would be compatible with both the high phylogenetic proximity between *At4CL1* and *At4CL2*, as opposed to the much larger differences among all other combinations, and the structural and putative functional divergence of *At4CL2/4* and *At4CL3* before the evolution of the

*At4CL1* and *At4CL2* pair. Fig. 4 furthermore implies (i) that the separation into classes I and II predates the divergence of mono- and dicotyledonous plants, likely the outcome of an ancient polyploidization (56), (ii) that the two sinapate-activating isoenzymes, *At4CL4* and *Gm4CL1*, might have evolved independently, a possible explanation for the rare occurrence of this type of 4CL, and (iii) that *At4CL1* and *At4CL2* result from a single polyploidy event dating between the *A. thaliana*-*Brassica* split and the *A. thaliana*-cotton split (57).

In conclusion, the small *At4CL* family is characterized by a highly diversified substructure with a broad spectrum of enzyme-kinetic properties *in vitro* (18) and equally diversified expression patterns *in vivo* (J. Ehling and E. Kombrink, personal communication). This complexity of structural and putative functional features at a central position in phenylpropanoid biosynthesis renders the *At4CL* family ideally suited for more detailed studies on the individual metabolic roles of the four isoenzymes *in vivo*. In particular, elucidation of the mechanisms of substrate channeling into the numerous major and minor branches of phenylpropanoid metabolism is likely to yield important clues for an understanding of the large species-specific diversity and the functional significance of the myriad phenylalanine-derived secondary compounds occurring in all higher plants.

We thank Dr. Jürgen Ehling and Prof. Carl Douglas (University of British Columbia) for critical comments and reading of the manuscript, and Drs. Jürgen Ehling and Erich Kombrink (Max Planck Institute for Plant Breeding Research, Cologne, Germany) for communicating unpublished results.

- Kumar, A. & Ellis, B. E. (2003) *Plant Mol. Biol.* **51**, 327–340.
- Wallis, P. J. & Rhodes, M. J. C. (1977) *Phytochemistry* **16**, 1891–1894.
- Knobloch, K. H. & Hahlbrock, K. (1977) *Arch. Biochem. Biophys.* **184**, 237–248.
- Hu, W. J., Kawaoka, A., Tsai, C. J., Lung, J., Osakabe, K., Ebinuma, H. & Chiang, V. L. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 5407–5412.
- Ranjeva, R., Boudet, A. M. & Faggion, R. (1976) *Biochimie* **58**, 1255–1262.
- Lozoya, E., Hoffmann, H., Douglas, C., Schulz, W., Scheel, D. & Hahlbrock, K. (1988) *Eur. J. Biochem.* **176**, 661–667.
- Allina, S. M., Pri-Hadash, A., Theilmann, D. A., Ellis, B. E. & Douglas, C. J. (1998) *Plant Physiol.* **116**, 743–754.
- Knobloch, K. H. & Hahlbrock, K. (1975) *Eur. J. Biochem.* **52**, 311–320.
- Kutsuki, H., Shimada, M. & Higuchi, T. (1982) *Phytochemistry* **21**, 267–271.
- Grand, C., Boudet, A. & Boudet, A. (1983) *Planta* **158**, 225–229.
- Lindermayr, C., Mollers, B., Fliegmann, J., Uhlmann, A., Lottspeich, F., Meimberg, H. & Ebel, J. (2002) *Eur. J. Biochem.* **269**, 1304–1315.
- Stuible, H. P. & Kombrink, E. (2001) *J. Biol. Chem.* **276**, 26893–26897.
- Schneider, K., Hovel, K., Witzel, K., Hamberger, B., Schomburg, D., Kombrink, E. & Stuible, H. P. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 8601–8606.
- Hahlbrock, K., Bednarek, P., Ciolkowski, I., Hamberger, B., Heise, A., Liedgens, H., Logemann, E., Nürnberger, T., Schmelzer, E., Somssich, I. E. & Tan, J. (2003) *Proc. Natl. Acad. Sci. USA* **100**, Suppl. 2, 14569–14576.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
- Heise, A., Lippok, B., Kirsch, C. & Hahlbrock, K. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 9049–9054.
- Stuible, H., Büttner, D., Ehling, J., Hahlbrock, K. & Kombrink, E. (2000) *FEBS Lett.* **467**, 117–122.
- Ehling, J., Büttner, D., Wang, Q., Douglas, C. J., Somssich, I. E. & Kombrink, E. (1999) *Plant J.* **19**, 9–20.
- Ragg, H., Kuhn, D. N. & Hahlbrock, K. (1981) *J. Biol. Chem.* **256**, 10061–10065.
- Devereux, J., Haerberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
- Fitch, W. M. (1977) *Am. Nat.* **111**, 223.
- Swofford, D. L. & Olsen, G. J. (1990) in *Molecular Systematics*, eds. Hillis, D. M. & Moritz, C. (Sinauer, Sunderland, MA), pp. 411–501.
- Felsenstein, J. (1985) *Evolution* **39**, 783–791.
- Higo, K., Ugawa, Y., Iwamoto, M. & Korenaga, T. (1999) *Nucleic Acids Res.* **27**, 297–300.
- Eulgem, T., Rushton, P. J., Robatzek, S. & Somssich, I. E. (2000) *Trends Plant Sci.* **5**, 199–206.
- Cukovic, D., Ehling, J., VanZiffle, J. A. & Douglas, C. J. (2001) *Biol. Chem.* **382**, 645–654.
- Lois, R., Dietrich, A., Hahlbrock, K. & Schulz, W. (1989) *EMBO J.* **8**, 1641–1648.
- Logemann, E., Parniske, M. & Hahlbrock, K. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 5905–5909.
- Arabidopsis Genome Initiative (2000) *Nature* **408**, 796–815.
- Becker-Andre, M., Schulze-Lefert, P. & Hahlbrock, K. (1991) *J. Biol. Chem.* **266**, 8551–8559.
- Raes, J., Rohde, A., Christensen, J. H., Van De Peer, Y. & Boerjan, W. (2003) *Plant Physiol.* **133**, 1051–1071.
- Harding, S. A., Leshkevich, J., Chiang, V. L. & Tsai, C. J. (2002) *Plant Physiol.* **128**, 428–438.
- Lee, D. & Douglas, C. J. (1996) *Plant Physiol.* **112**, 193–205.
- Zubieta, C., Kota, P., Ferrer, J. L., Dixon, R. A. & Noel, J. P. (2002) *Plant Cell* **14**, 1265–1277.
- Anterola, A. M. & Lewis, N. G. (2002) *Phytochemistry* **61**, 221–294.
- Dixon, R. A., Chen, F., Guo, D. & Parvathi, K. (2001) *Phytochemistry* **57**, 1069–1084.
- Koopmann, E., Logemann, E. & Hahlbrock, K. (1999) *Plant Physiol.* **119**, 49–56.
- Rasmussen, S. & Dixon, R. A. (1999) *Plant Cell* **11**, 1537–1551.
- Burbulis, I. E. & Winkel-Shirley, B. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 12929–12934.
- Kajita, S., Katayama, Y. & Omori, S. (1996) *Plant Cell Physiol.* **37**, 957–965.
- Lee, D., Meyer, K., Chapple, C. & Douglas, C. J. (1997) *Plant Cell* **9**, 1985–1998.
- Schoch, G., Goepfert, S., Morant, M., Hehn, A., Meyer, D., Ullmann, P. & Werck-Reichhart, D. (2001) *J. Biol. Chem.* **276**, 36566–36574.
- Li, L., Popko, J. L., Umezawa, T. & Chiang, V. L. (2000) *J. Biol. Chem.* **275**, 6537–6545.
- Humphreys, J. M., Hemm, M. R. & Chapple, C. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 10045–10050.
- Terashima, N., Fukushima, K., Tsuchiya, S. & Takabe, K. (1986) *J. Wood Chem. Technol.* **6**, 495–504.
- Kirst, M., Johnson, A. F., Baucom, C., Ulrich, E., Hubbard, K., Staggs, R., Paule, C., Retzel, E., Whetten, R. & Sederoff, R. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 7383–7388.
- Goff, S. A., Ricke, D., Lan, T. H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., et al. (2002) *Science* **296**, 92–100.
- Yu, J., Hu, S., Wang, J., Wong, G. K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., et al. (2002) *Science* **296**, 79–92.

49. Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., *et al.* (2003) *Science* **301**, 376–379.
50. Grimmig, B. & Matern, U. (1997) *Plant Mol. Biol.* **33**, 323–341.
51. Hahlbrock, K. (1976) *Eur. J. Biochem.* **63**, 137–145.
52. Hahlbrock, K., Scheel, D., Logemann, E., Nurnberger, T., Parniske, M., Reinold, S., Sacks, W. R. & Schmelzer, E. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 4150–4157.
53. Batz, O., Logemann, E., Reinold, S. & Hahlbrock, K. (1998) *Biol. Chem.* **379**, 1127–1135.
54. Hahlbrock, K., Ebel, J., Ortmann, R., Sutter, A., Wellmann, E. & Grisebach, H. (1971) *Biochim. Biophys. Acta* **244**, 7–15.
55. Hahlbrock, K. & Grisebach, H. (1970) *FEBS Lett.* **11**, 62–64.
56. Bowers, J. E., Chapman, B. A., Rong, J. & Paterson, A. H. (2003) *Nature* **422**, 433–438.
57. Blanc, G., Hokamp, K. & Wolfe, K. H. (2003) *Genome Res.* **13**, 137–144.