Phytochemistry 69 (2008) 2449–2456

Contents lists available at ScienceDirect

Phytochemistry

journal homepage: www.elsevier.com/locate/phytochem

Identification of a *4-coumarate:CoA ligase* gene family in the moss, *Physcomitrella patens* ☆

Martina V. Silber [a], Harald Meimberg [b], Jürgen Ebel [a,*]

[a] *Department Biologie I – Botanik, Ludwig-Maximilians-Universität, Menzinger Strasse 67, D-80638 München, Germany*
[b] *CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, R. Monte-Crasto, 4485-661 Vairão, Portugal*

ARTICLE INFO

Article history:
Received 11 October 2007
Received in revised form 7 May 2008
Available online 21 August 2008

Keywords:
Physcomitrella patens
Funariaceae
Moss
Enzyme evolution
Phenylpropanoid metabolism
4-Coumarate:CoA ligase
Multigene family

ABSTRACT

Since the early evolution of land plants from primitive green algae, phenylpropanoid compounds have played an important role. In the biosynthesis of phenylpropanoids, 4-coumarate:CoA ligase (4CL; EC 6.2.1.12) has a pivotal role at the divergence point from general phenylpropanoid metabolism to several major branch pathways. Although higher plant 4CLs have been extensively studied, little information is available on the enzymes from bryophytes. In *Physcomitrella patens*, we have identified a *4CL* gene family consisting of four members, taking advantage of the available EST sequences and a draft sequence of the *P. patens* genome. The encoded proteins of three of the genes display similar substrate utilization profiles with highest catalytic efficiency towards 4-coumarate. Interestingly, the efficiency with cinnamate as substrate is in the same range as with caffeate and ferulate. The deduced proteins of the four genes share sequence identities between 78% and 86%. The intron/exon structures are pair wise similar. *Pp4CL2* and *Pp4CL3* each consists of four exons and three introns, whereas *Pp4CL1* and *Pp4CL4* are characterized each by five exons and four introns. *Pp4CL1*, *Pp4CL2* and *Pp4CL3* are expressed in both gametophore and protonema tissue of *P. patens*, unlike *Pp4CL4* whose expression could not be demonstrated under the conditions employed. Phylogenetic analysis suggests an early evolutionary divergence of *Pp4CL* gene family members. Using *Streptomyces coelicolor* cinnamate:CoA ligase (*Sc*CCL) as an outgroup, the *P. patens* 4CLs are clearly separated from the spermatophyte proteins, but are intercalated between the angiosperm 4CL class I and class II. A comparison of three *P. patens* subspecies from diverse geographical locations shows high sequence identities for the four 4CL isoforms.

1. Introduction

The bryophytes contain a large number of secondary compounds, several of which are bioactive constituents which protect the organisms from herbivores, fungi, bacteria, slugs, and snails. Among the secondary constituents, major phenylpropanoid-derived classes of compounds are flavonoids, bibenzyls, and bis(bibenzyl)s with widespread occurrence in liverworts and mosses (Markham, 1988; Asakawa, 2001). Central to many of phenylpropanoid biosynthetic pathways is the activation of 4-coumarate (**2** in Fig. 1) and a number of related substrates to the corresponding CoA thioesters catalyzed by 4-coumarate:CoA ligase (4CL; EC 6.2.1.12). The CoA thioester products constitute common C_9 building units for proposed specific branch pathways subsequent to general phenylpropanoid metabolism.

Because of the importance of phenylpropanoid-derived compounds in plants, 4CL has been the subject of extensive study for many years, mainly in higher plants. It has been shown to occur in the form of multiple isoenzymes with either similar (e.g. Lozoya et al., 1988) or distinct (e.g. Lindermayr et al., 2002) substrate affinities. More recently, a gene encoding a 4CL-like enzyme with more than 40% identity in amino acid sequence to higher plant 4CLs was found in the genome of the filamentous bacterium, *Streptomyces coelicolor* A3(2) (Kaneko et al., 2003). The recombinant enzyme showed distinct 4CL activity with highest catalytic efficiency towards cinnamate (**1**). However, only limited information is available on 4CL and other enzymes of phenylpropanoid biosynthesis in bryophytes. Recently, a multigene family of chalcone synthase, the enzyme catalyzing the committed step of flavonoid biosynthesis, was characterized in the moss, *Physcomitrella patens* (Jiang et al., 2006).

Abbreviations: 4CL, 4-coumarate:CoA ligase; EST, expressed sequence tag

☆ Data deposition: EU167552 (*Pp*4CL1, *P. patens* ssp. *patens*), EU167553 (*Pp*4CL2, *P. patens* ssp. *patens*), EU167554 (*Pp*4CL3, *P. patens* ssp. *patens*), EU167555 (*Pp*4CL4, *P. patens* ssp. *patens*), EU179312 (*Pp*4CL1, *P. patens* ssp. *californica*), EU179313 (*Pp*4CL2, *P. patens* ssp. *californica*), EU179314 (*Pp*4CL3, *P. patens* ssp. *californica*), EU179315 (*Pp*4CL4, *P. patens* ssp. *californica*), EU180599 (*Pp*4CL1, *P. patens* ssp. *magdalenae*), EU180600 (*Pp*4CL2, *P. patens* ssp. *magdalenae*), EU180601 (*Pp*4CL3, *P. patens* ssp. *magdalenae*), EU180602 (*Pp*4CL4, *P. patens* ssp. *magdalenae*).

* Corresponding author. Tel.: +49 (0)89 17861285; fax: +49 (0)89 1782274.
E-mail address: j.ebel@lrz.uni-muenchen.de (J. Ebel).

0031-9422/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2008.06.014

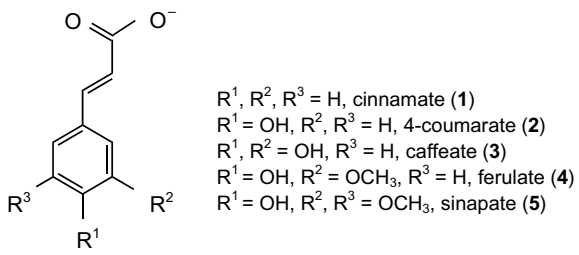

Fig. 1. Chemical structures of the five naturally occurring 4CL substrates investigated in this study.

P. patens has drawn much attention because of its evolutionary position among land plants and of its exceptional genetic properties (Schaefer and Zrÿd, 2001). Thus, *P. patens* may represent an ideal case for studies on the evolution of metabolic pathways including phenylpropanoid biosynthesis. As a result of multiple sequencing projects, nucleic acid sequences of >100,000 expressed sequence tags (ESTs) and >16,000 annotated transcripts are available (Nishiyama et al., 2003; Lang et al., 2005). In addition, the first draft sequence of the *P. patens* genome has recently been released (http://shake.jgi-psf.org/Phypa1/Phypa1.home.html; Rensing et al., 2008). In this study, by exploiting the various public resources, the existence of a family of four closely related 4CL genes in the moss, *P. patens*, was demonstrated. For three of the encoded proteins, we verified the identity as 4CL isoenzymes and demonstrated that the respective genes are expressed in moss tissues of different developmental stages. Phylogenetic reconstructions indicated that the four *Pp*4CL proteins are closely related and may have resulted from comparably recent gene duplication events. They are forming together with the gymnosperm proteins the neighbour group of the angiosperm 4CL proteins of class II. This indicates that the split between 4CL class I and II occurred very early in the evolution of green land plants.

2. Results

2.1. Isolation of cDNAs and genes, heterologous expression, and biochemical characterization of Pp4CL isoforms

Analysis of the publicly available *P. patens* ESTs (*Physcomitrella* GenBank EST databases) was performed using the following selective criteria for the identification of putative *4CL* sequences: (i) the presence and identity of *4CL*-specific nucleotide sequences which encode peptides representing the putative AMP-binding domain (Box I) and the conserved GEICIRG motif (Box II) (Stuible and Kombrink, 2001) within candidate proteins; (ii) sequences with significant homology to the soybean *Gm*4CL1 cDNA. The search revealed the presence of one putative *Pp*4CL cDNA which was designated *Pp4CL1*. The presence of putative *Pp4CL2–4* genes was indicated by either screening of a bacterial artificial chromosome (BAC) library (Centre of Plant Sciences, LIBA, Leeds University, UK) (*Pp4CL2*) or analysis of the publicly accessible *P. patens* genome sequence data (Joint Genome Institute of US Department of Energy, USA) (*Pp4CL3* and *Pp4CL4*). Full-length cDNA clones for *Pp*4CL1 and *Pp*4CL3 were obtained from RIKEN BioResource Center, Japan (Nishiyama et al., 2003), whereas the *Pp*4CL2 cDNA was cloned from total RNA isolated from *P. patens* gametophores by RT-PCR using gene specific primers as detailed in Table 1. No EST clone existed for *Pp*4CL4 and no expression of this gene could be detected under the conditions used in this study (see below). We therefore did not attempt to isolate a cDNA representing this isoform. Cloning of the *Pp4CL1–4* genes was performed by amplifying genomic DNA products using Pfu DNA polymerase and pairs of oligonucleotide primers as listed in Table 1. The enzymes encoded by *Pp*4CL1–3 cDNAs were synthesized as N-terminally His_6-tagged proteins in *Escherichia coli* and their immunogenic cross-reactivity was demonstrated by using an antiserum raised against parsley 4CL (Ragg et al., 1981; data not shown). After immobilized metal affinity chromatography, the three recombinant proteins revealed the relative molecular masses in SDS/polyacrylamide gels of 65 kDa for *Pp*4CL1 and *Pp*4CL2, respectively, and 64 kDa for *Pp*4CL3.

Table 1
Gene specific sequences of oligonucleotides to amplify putative *Physcomitrella patens 4CL1–4* genes and transcripts

Gene	Designation	Sequence
Pp4CL1	Pp4CL1_5′UTR	5′-TATTGTAGAGTCCGATAGCGTC
	Pp4CL1_3′UTR	5′-TAAACCCTAAGTTGCAGGATCTG
	Pp4CL1RT_PCRfor	5′-TGGAATGACAGAAGCTGGTCC
	Pp4CL1RT_PCRrev	5′-CCTAAGTTGCAGGATCTGGCTC
Pp4CL2	Pp4CL2_5′UTR	5′-GGAAGCAAGAGTTTCGGAG
	Pp4CL2_3′UTR	5′-ACATTGTGGGAGGATACCGC
	Pp4CL2RT_PCRfor	5′-AGCATCAGGATGGTGATGTCCG
	Pp4CL2RT_PCRrev	5′-ACATTGTGGGAGGATACCGC
Pp4CL3	Pp4CL3_5′UTR	5′-ACAGATCGGTGAGCACATTGG
	Pp4CL3_3′UTR	5′-TGTTTCAGGAATGGATGCGTC
	Pp4CL3RT_PCRfor	5′-CCAGGATCTTGTGGCACTGTC
	Pp4CL3RT_PCRrev	5′-AGGCTCACGATTCCATCCGAC
Pp4CL4	Pp4CL4_5′UTR	5′-TAGGTTCGGTAATCCTCCACC
	Pp4CL4_3′UTR	5′-TGGTATGGTTCCCGTATCGAG
	Pp4CL4RT_PCRfor	5′-TGGCAAGGAGCTTGAGGATGC
	Pp4CL4RT_PCRrev	5′-TCTTGAAACCTACGGATGAAGACTC

The enzyme-kinetic analyses of the purified recombinant proteins (Table 2) verified their biochemical function as bona fide 4CLs. All of the enzymes displayed Michaelis–Menten kinetics, however with the inherent problem of instability of the catalytic activity of the affinity-purified proteins. We therefore decided to estimate K_m and relative V_{max} values which limits to some extent the comparison of the data for the different iosoenzymes in terms of their molar activities. *Pp*4CL1–3 efficiently converted cinnamate (**1**), 4-coumarate (**2**), caffeate (**3**), and ferulate (**4**) to the corresponding CoA esters. According to the K_m and V_{max}/K_m values, cinnamate (**1**), which is poorly activated by many higher plant 4CLs, was converted at a rate similar to that for caffeate (**3**) and ferulate (**4**), whereas 4-coumarate (**2**) proved to be the most efficient of the four tested naturally occurring 4CL substrates (Table 2). Sinapate (**5**) was not accepted as a substrate under the experimental conditions, in line with most previously characterized 4CLs from a wide range of higher plant species.

2.2. Comparative analysis of Pp4CL gene family

The nucleotide and amino acid sequence similarities within the *Pp4CL* gene family share a percentage of identity between 78% and 86% (amino acid sequences) and 77–83% (nucleotide sequences of the central part of the coding region of the genes omitting about 180 base pairs of the N-terminus). The overall identity of the *Pp*4CL isoforms versus 4CLs from higher plants is about 62%. When compared with 4CLs from higher plants, *Pp*4CL1–4 contained an extension of about 30 amino acids at the N-terminal part of the proteins. At present it is not known whether this N-terminal extension has any functional impact on the *Pp*4CLs.

Fig. 2 illustrates the exon/intron structures of the four *Pp4CL* genes. *Pp4CL2* and *Pp4CL3* each contain four exons and three introns of similar sizes, respectively. For both, *Pp4CL1* and *Pp4CL4*, the major difference to the former pair is represented by an additional intron in the first exon.

To describe the expression patterns of each member of the *Pp4CL* gene family, specific oligonucleotides were designed from the particular 3′ untranslated and 5′ coding region of the cDNAs.

Table 2
Kinetic properties of *P. patens* *Pp*4CL1–3 *in vitro*

Substrate	K_m (μM)			Specific activity relative V_{max}			Relative V_{max}/K_m (μM^{-1})		
	*Pp*4CL1	*Pp*4CL2	*Pp*4CL3	*Pp*4CL1	*Pp*4CL2	*Pp*4CL3	*Pp*4CL1	*Pp*4CL2	*Pp*4CL3
Cinnamate (**1**)	50	66	64	61	59	57	1.19	0.89	0.89
4-Coumarate (**2**)	4	16	29	100	100	100	25.97	6.25	3.5
Caffeate (**3**)	25	45	725	24	46	190	0.9	1.0	0.26
Ferulate (**4**)	187	73	800	114	108	195	0.61	1.4	0.25

All data are mean values from three independent experiments with affinity-purified proteins. Relative V_{max} values were obtained by setting V_{max} of 4-coumarate for each isoform to 100%. Sinapate (**5**) was not accepted as a substrate. See Fig. 1 for chemical structures of substrates.

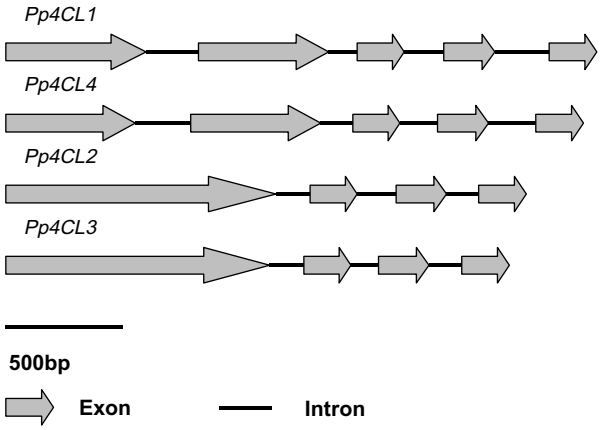

Fig. 2. Comparison of exon/intron structures of bona fide *Pp4CL1–3* and putative *Pp4CL4* of *Physcomitrella patens*.

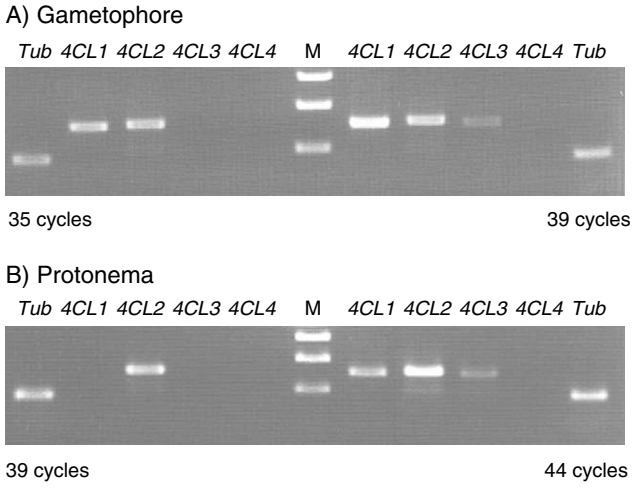

Fig. 3. Expression patterns of bona fide and putative *4CL* genes in two developmental stages of *P. patens*. Total RNA isolated from 2-week-old gametophore and 4-day-old protonema tissue was used for cDNA synthesis. cDNA was analyzed by semi-quantitative PCR using gene-specific primers and 35 (39) amplification cycles for gametophore (A) and 39 (44) amplification cycles for protonema (B). Lane M, size marker; lane *Tub*, *P. patens* β-tubulin gene.

As detailed under Experimental, isogene specificity of each primer pair was proved by PCR using cloned DNA as template. The gene for β-tubulin, earlier shown to be constitutively expressed (Jost et al., 2004), was used as a control. As illustrated in Fig. 3, the genes *Pp4CL1–3*, but not *Pp4CL4*, were transcribed in both gametophore and protonema tissue under the experimental conditions employed. To enhance the detection limit for low amounts of *Pp4CL4* transcripts, the number of amplification cycles was further increased to 47 and the amount of reaction mixture applied to electrophoresis was enhanced four-fold compared to the standard conditions (Fig. 3). This modification again did not result in the detection of any *Pp4CL4* transcripts. Using semi-quantitative PCR conditions, an apparent gradual strength of expression was evident: the highest level was visible for *Pp4CL2* and the lowest for *Pp4CL3* in both tissues. Similar results were obtained when analyzing RNA from three independent preparations.

2.3. Phylogenetic analysis of the plant 4CL family

Phylogenetic reconstructions were performed to evaluate the relationship of the different *P. patens* 4CLs with all available amino acid sequences of the bona fide plant 4CL protein family (Fig. 4). The *Pp*4CL isoenzymes form a highly supported monophyletic group and are thus separated from higher plant isoforms. They are positioned as a neighbour group to the gymnosperm sequences, the pine isoforms *Pt*4CL1 and *Pt*4CL2. In accordance with earlier studies (Ehlting et al., 1999; Lindermayr et al., 2002), angiosperm 4CLs are divided in two groups, designated 4CL class I and class II as defined earlier (Ehlting et al., 1999). Using the microbial enzyme, *S. coelicolor* A3(2) cinnamate:CoA ligase (*Sc*CCL), as outgroup, 4CL isoenzymes of class II appear more closely related to the *Pp*4CL isoforms than to the 4CL isoforms of class I.

Although the overall sequence similarity between all isolated *Pp4CL* family members at the nucleotide and protein level is relatively high (78–86% of amino acid sequence identity; 77–83% of nucleotide sequence identity among the *Pp*4CL1–4 isoforms), the most recent gene duplication event might have occurred between *Pp4CL1* and *Pp4CL4*. In addition, genome duplication has been proposed to have occurred in *P. patens* between 30 and 60 million years ago (Rensing et al., 2007). Both events thus might in part explain the versatility of metabolism including secondary metabolism, as described for *P. patens* and other mosses, in comparison to other land plants (Rensing et al., 2007).

Three *P. patens* subspecies have been analyzed to detect evolutionary relationships of 4CL within the species. In particular, subspecies *patens* (*ecotype Gransden 2004* from the UK which has been the basis for the EST and the genome sequencing projects), ssp. *magdalenae* (*ecotype Bisoke*, Africa), and ssp. *californica* (*ecotype Okayama*, Japan) were used for comparison. Using sequence information of the isolated *4CL* genes of *P. patens* ssp. *patens*, the two sets of *4CL* genes were isolated by PCR from genomic DNA of the additional subspecies.

P. patens ssp. *magdalenae* as well as *P. patens* ssp. *californica* contain a family of four *4CL* genes with an exon/intron structure comparable to that of *P. patens* ssp. *patens*. The lengths of the exons were identical except for exon one of *4CL2* from *P. patens* ssp. *magdalenae* being six nucleotides shorter than the respective exons from the other two subspecies, whereas the size of the introns appeared to be more variable. The percentages of identity of the nucleotide and amino acid sequences within each 4CL family of the additional subspecies closely matched those of *P. patens* ssp. *patens*. A comparison of the amino acid similarities for each of

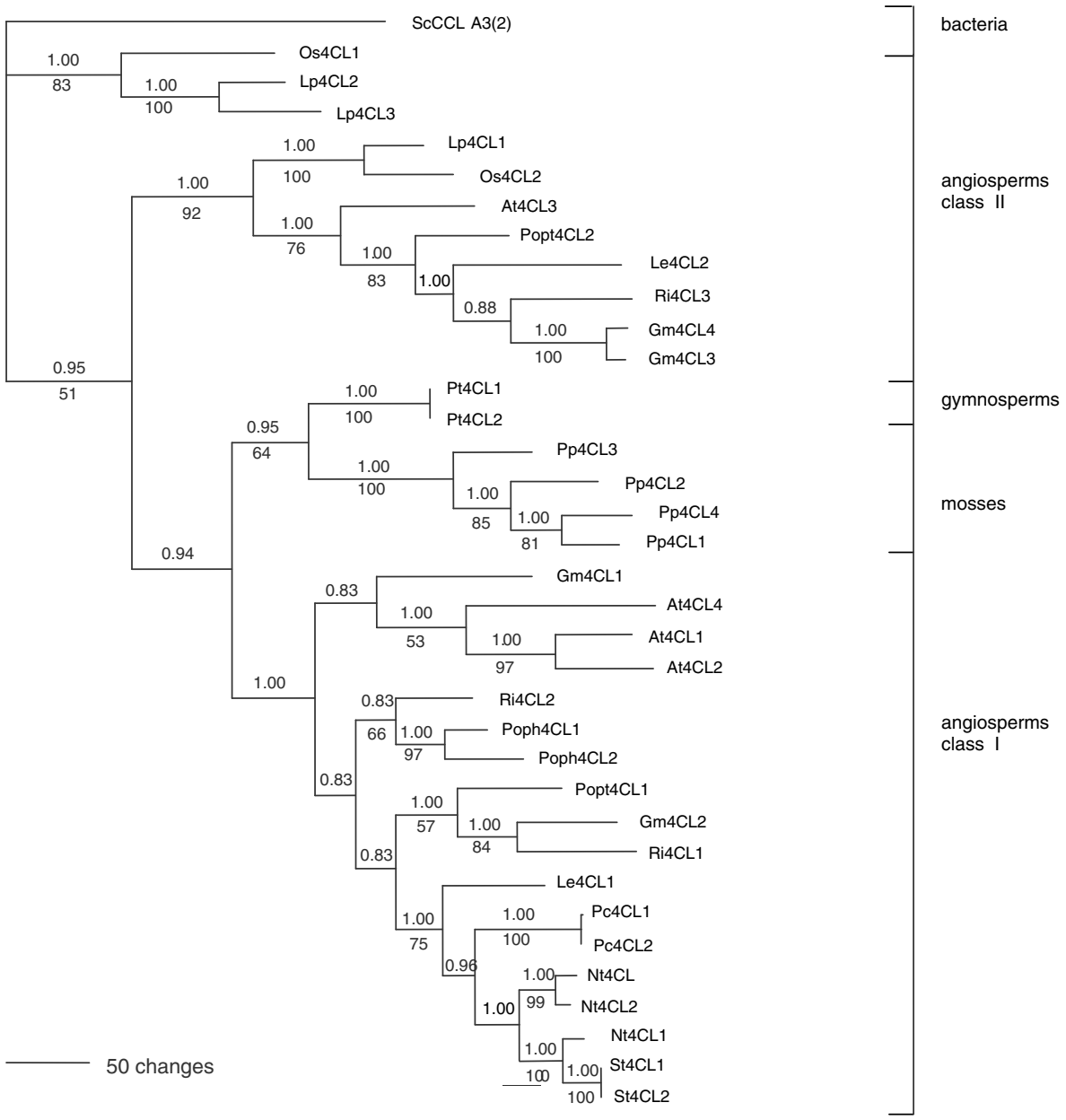

Fig. 4. Phylogram of published members of the bona fide 4CL family from higher plants in comparison to 4CL isoforms, *Pp*4CL1–4, determined for the moss *P. patens*. Bayesian analysis was in general accordance to a maximum parsimony analysis. Bayesian posterior probabilities are indicated above branches, corresponding bootstrap values of the maximum parsimony analysis from 1000 replicates are given below branches. As outgroup, the bacterial cinnamate:CoA ligase (*Sc*CCL) from *Streptomyces coelicolor* A3(2) was chosen.

the 4CL isoforms from the three ecotypes revealed a percentage of identity between 95 (*Pp*4CL4) and 97 (*Pp*4CL1).

The phylogenetic analysis of the 4CL isoforms from *P. patens* ssp. *patens*, *magdalenae* and *californica* (data not shown) supported (i) a relatively large genetic distance between the Japanese (e.g. ssp. *californica*) and the European (e.g. ssp. *patens*) subspecies based on internal transcribed spacer regions as concluded earlier (von Stackelberg and Reski, unpublished results), (ii) a proposed closer relationship between the Japanese and the African subspecies (e.g. ssp. *magdalenae*), and in addition (iii) the largest divergence of 4CL4 within the *4CL* gene family of *P. patens*.

3. Discussion

The acyl-activating capacity of CoA ligases, which belong to the superfamily of adenylate-forming enzymes, is believed to constitute a central function in the biosynthesis of a host of both primary and secondary compounds in various pathways. Among these, the biosynthesis of many phenylpropanoid compounds in green land plants involves the activation of 4-coumarate and a number of structurally related substrates to the corresponding CoA thioesters which is catalyzed by 4CL. As 4CL has been studied until present mainly in higher plants, the prime objective of this study was to extend the molecular survey of 4CL to bryophytes.

The moss 4CL family reported here shows similar enzymatic characteristics when compared to higher plant 4CLs. Considering the similarities in substrate specificity and in the protein sequence, the *Pp*4CL isoforms are representing one of the evolutionary oldest 4CL families of land plants and are linking the phylogeny of *S. coelicolor Sc*CCL to 4CL isoforms of higher plants. The close relationship of the *Pp*4CL isoforms to the higher plant

4CL family in the phylogenetic tree together with the catalytic preference within the group of 4-coumarate (**2**) and related substrates of the *Pp*4CL isoforms underlines the importance of 4CL as a central enzyme of phenylpropanoid metabolism of all plants. The position of *Pp*4CL isoforms as a neighbour group to the angiosperm 4CL isoforms of class II indicates that the split between class I and class II isoforms occurred at a very early stage in the evolution of land plants, before the divergence of the lineages that lead to mosses on the one hand and to higher plants on the other. This illustrates further the high importance of phenylpropanoid pathways for plants.

*Pp*4CL likely consists of a family of isoforms which exhibits largely similar structures and enzyme-kinetic properties. In this regard, *Pp*4CL resembles *Solanum tuberosum St*4CL, *Petroselinum crispum Pc*4CL, and *Nicotiana tabacum Nt*4CL (Lozoya et al., 1988; Becker-Andre et al., 1991; Lee and Douglas, 1996). The opposite is exemplified by 4CL families which exhibit large structural, substrate specificity, and apparent evolutionary diversity, such as *Arabidopsis thaliana At*4CL, *Glycine max Gm*4CL, *Lithospermum erythrorhizon Le*4CL, *Populus tremuloides Popt*4CL, and *Rubus idaeus Ri*4CL (Yazaki et al., 1995; Hu et al., 1998; Lindermayr et al., 2002; Kumar and Ellis, 2003; Hamberger and Hahlbrock, 2004; Costa et al., 2005). To further advance our knowledge of the molecular evolution of the 4CL family, the analyses have to be extended to 4CL members from the pteridophytes and microorganisms.

The overall amino acid sequence identity between the four *Pp*4CL isoforms and higher plant 4CLs is about 62%, as opposed to 42–43% when compared with the *Sc*CCL from *S. coelicolor* A3(2) whose catalytic function has been described as cinnamate:CoA ligase (Kaneko et al., 2003). Other 4CL-like sequences from higher plants have been identified. For example, the genome of *A. thaliana* has 14 genes, annotated as putative 4CL isoforms (Costa et al., 2003). One of the closest relatives to *At*4CL1–4, *At*1g20510, has recently been identified as 12-oxo-phytodienoic acid (OPDA):CoA ligase (OPCL1) and proposed to be involved in the biosynthesis of jasmonic acid in *Arabidopsis* (Koo et al., 2006). OPCL1 is a member of clade V of the phylogenetic tree of the 4CL-like subfamily of acyl-activating enzymes in *Arabidopsis*, all of which are characterized by a peroxisomal targeting signal (Koo et al., 2006). Two further members of this group were previously shown to encode peroxisomal enzymes that activate jasmonic acid precursors in vitro as well (Schneider et al., 2005). The sequence relationship between *Pp*4CLs and the clade V member, OPCL1 (36–37% identity with *Pp*4CL1–4), appears more remote than between *Pp*4CLs and *S. coelicolor* 4CL (42–43%) which is in agreement with the divergent biochemical functions of the respective enzymes. As the homology criteria used for the identification of *Pp*4CLs likely precluded the identification of putative *Pp*4CL-like proteins, no conclusion can be drawn at present as to the existence and properties of the latter proteins in *P. patens*. For other 4CL-like homologues of higher plants, beyond database annotations, there is a need for demonstrating any specific enzyme function. To better understand the catalytic specificity of 4CLs and 4CL-like enzymes, however, specificity-determining amino acids of the substrate-binding pockets will have to be identified (Kaneko et al., 2003; Lindermayr et al., 2003; Schneider et al., 2003). Such studies could also reveal why *Pp*4CLs accept cinnamate much better than many higher plant 4CLs. Comprehensive in silico studies, in the same context, could identify functionally undefined 4CL-like proteins in *P. patens* involved in various acyl-activating reactions.

The exon/intron structures of *4CL* genes from higher plants, in most cases, have not been analyzed in great detail. The total number and relative position of the introns within the genes can vary to some extent, as found for example for the *At4CL* genes (three introns in *At4CL1*, six introns in *At4CL3*; see Hamberger and Hahlbrock, 2004) or for the *Gm4CL* genes (four introns in *Gm4CL1*, five introns in *Gm4CL4*; unpublished results). The position of the first intron in higher plant *4CLs*, however, appears to match the position of intron one of *Pp4CL2*. Whereas the exon/intron organisation of the known four *4CL* genes in soybean is more closely related to that of *Pp4CL2* and *Pp4CL3*, that of *At4CL3* is structurally more similar to *Pp4CL1* and *Pp4CL4*.

Three of the four *P. patens 4CL* genes were expressed under the experimental conditions used in our studies and in those employed for collecting ESTs. The encoded proteins displayed a substrate utilization capacity similar to higher plant 4CLs, however with the added increase in activity towards cinnamate. Isoforms 1–3 thus represent bona fide 4CLs. At least one biochemical function of *Pp*4CL is obvious, namely its provision of activated C_9 precursors for flavonoid biosynthesis in *P. patens* (Jiang et al., 2006). Comparative genome analyses demonstrated the presence in *P. patens* (Rensing et al., 2008) of orthologues of phenylalanine ammonia-lyase and of CYP73, cinnamate 4-hydroxylase, which, together with 4CL, comprise enzymes of core phenylpropanoid metabolism, and of CYP98, coumaroyl-shikimate/quinate 3′-hydroxylase, an additional member of the pathway assumed to be the major 3-hydroxylase in phenylpropanoid metabolism of higher plants (Ehlting et al., 2006). More refined functional assignments of the seemingly similar *Pp*4CL isoforms may require comparative studies of the family members, including gene-specific knock-out mutations combined with histochemical, physiological, and biochemical analyses. In addition, some of the higher plant *4CL* genes are constitutively expressed while others are transcriptionally induced or repressed in response to environmental factors such as UV light and pathogen attack. Little is known at present concerning the contribution of phenylpropanoid compounds in the adaptation of mosses to environmental challenges. It would be of particular interest to learn about the regulation of *Pp4CLs* in response to external factors and to unravel the evolutionary relationships of regulation and possible functional divergence of the isoenzymes.

4. Conclusions

The *Pp*4CL family of *Physcomitrella* was characterized and its enzymatic properties in vitro were determined to be similar to 4CL families of those higher plants that contain isoenzymes of largely similar structures and catalytic properties. The substrate preferences suggested that the *Pp*4CL family members play a central role in phenylpropanoid biosynthesis of the moss comparable to that of higher plant 4CLs. Of the four isolated genes, three were expressed under the experimental growth conditions used in the studies. Future investigations are required to yield clues on the regulation of the *Pp*4CL family members and on their contribution to growth and adaptation to environmental challenges of the moss.

5. Experimental

5.1. Plant material and growth conditions

P. patens (Hedw.) B.S.G., kindly provided by Drs. R. Reski and M. von Stackelberg, was grown under sterile conditions in minimal medium (Ashton et al., 1979), supplemented with 0.5 g/l ammonium tartrate and 5 g/l glucose. Protonema tissues were propagated by fragmentation of gametophytes with a blender (MICCRA D-8Si, ART-Labortechnik, Müllheim, Germany) and subcultured weekly in 500 ml Erlenmeyer flasks containing 200 ml of medium. Gametophore tissues, grown on agar plates (0.7%, w/v; agar A-7921 from SIGMA), were subcultured at 4-week intervals. Plants were grown at 16 h light/day (Philips TL-D 36W/33; photon fluence of ca. 40 μmol s^{-1} m^{-2}) and at 25 °C.

5.2. Data retrieval

Similarity searches were performed by using the coding region of the soybean *Gm*4CL1 cDNA (Genbank accession no. AF279267) against the *Physcomitrella* GenBank EST database (http://moss.nibb.ac.jp/), allowing the retrieval of a candidate *Pp4CL* gene (GenBank accession no. BI437532) which was designated *Pp4CL1*. A full-length cDNA clone of *Pp*4CL1 (pdp11012) was obtained from RIKEN BioResource Center, 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan (Nishiyama et al., 2003).

Pp4CL3 and *Pp4CL4* were identified within the Community Sequencing Program at the Joint Genome Institute of US Department of Energy and made accessible for BLAST searches (Altschul et al., 1997) through PHYSCObase (http://moss.nibb.ac.jp) and the *Pp*4CL1 cDNA as query. A full-length *Pp*4CL3 cDNA clone (pdp35373) was obtained from RIKEN BioResource Center.

5.3. Molecular cloning of Pp4CL genes

Total cellular DNA was extracted using the cetyltrimethylammonium bromide (CTAB)-based method (Doyle and Doyle, 1990). Total RNA was prepared from protonema and gametophore tissues using the TriFast Reagent (peqlab, Biotechnologie GmbH, Erlangen, Germany) according to the instructions of the supplier. For reverse transcription, 1 μg of total RNA (gametophore tissue) and 2 μg of total RNA (protonema tissue) were incubated with RNasin and random primers (Invitrogen, Karlsruhe, Germany) for 7 min at 70 °C and then kept on ice. Reverse transcription was performed with M-MLV (Invitrogen, Karlsruhe, Germany) reverse transcriptase as directed by the manufacturer.

Pp4CL2 was identified by hybridizing a BAC filter (Centre of Plant Sciences, LIBA, Leeds University, UK) under high stringency conditions for 24 h at 60 °C using ^{32}P-labeld 400 bp of the C-terminal coding region of *Pp*4CL1. One clone was identified. Sequence information was obtained by direct sequencing of the corresponding BAC plasmid.

Genomic PCR products were amplified using Pfu DNA polymerase (Promega GmbH, Mannheim, Germany) with primers Pp4CL1_5′UTR/Pp4CL1_3′UTR (*Pp*4CL1), Pp4CL2_5′UTR/Pp4CL2_3′UTR (*Pp*4CL2), Pp4CL3_5′UTR/Pp4CL3_3′UTR (*Pp*4CL3), and Pp4CL4_5′UTR/Pp4CL4_3′UTR (*Pp*4CL4) (For a list of primers see Table 1. All oligonucleotides were obtained from MWG Biotech, Ebersberg, Germany) as follows: 1 cycle of 3 min at 94 °C, 34 cycles of 30 s at 94 °C, 1 min at 55 °C, and 5 min at 68 °C for amplification, and 1 cycle of 5 min at 68 °C for final extension. The resulting PCR products were cloned into the pDRIVE vector using the QIAGEN PCR Cloning Kit (Qiagen GmbH, Hilden, Germany).

5.4. Cloning of Pp4CL2 cDNA

To obtain a full-length cDNA of *Pp*4CL2, total RNA isolated from gametophore tissues was used for cDNA synthesis with the gene specific primer Pp4CL2_3′UTR. The same primer combined with primer Pp4CL2_5′UTR were used to amplify the coding region by PCR. The PCR product was inserted into the vector pDRIVE. Sequence data from this article have been deposited at GenBank under accession numbers EU167552 (*Pp*4CL1, *P. patens* ssp. *patens*), EU167553 (*Pp*4CL2, *P. patens* ssp. *patens*), EU167554 (*Pp*4CL3, *P. patens* ssp. *patens*), EU167555 (*Pp*4CL4, *P. patens* ssp. *patens*), EU179312 (*Pp*4CL1, *P. patens* ssp. *californica*), EU179313 (*Pp*4CL2, *P. patens* ssp. *californica*), EU179314 (*Pp*4CL3, *P. patens* ssp. *californica*), EU179315 (*Pp*4CL4, *P. patens* ssp. *californica*), EU180599 (*Pp*4CL1, *P. patens* ssp. *magdalenae*), EU180600 (*Pp*4CL2, *P. patens* ssp. *mahgdalenae*), EU180601 (*Pp*4CL3, *P. patens* ssp. *magdalenae*), EU180602 (*Pp*4CL4, *P. patens* ssp. *magdalenae*).

5.5. Expression of Pp4CL genes

For semi-quantitative PCR, gene-specific primers listed in Table 1 were used. Specific downstream/reverse primers were located within the corresponding 3′UTRs. Upstream/forward primers were designed specific to the coding region of *Pp4CL* genes (Table 1).

Length, G/C content, calculated annealing temperature and the expected lengths of the PCR products were in the same range for all primer pairs. Gene specificity for each primer pair was shown by PCR using cloned DNA as template (data not shown). At least three independent RNA preparations were analyzed. All PCR experiments were repeated at least once using 1.25 Units Taq recombinant DNA polymerase (QIAGEN GmbH, Hilden, Germany), 10 pmol of each primer and 100 ng (gametophore tissue) or 200 ng (protonema tissue) of initial total RNA in a volume of 50 μl for one reaction. Samples were loaded onto 1% (w/v) ethidium bromide-stained agarose gels in TAE buffer and visualized by illumination at $\lambda = 254$ nm. As an internal standard, *P. patens* β-tubulin gene 3 was used which was shown to be constitutively expressed in the tissues tested here (Jost et al., 2004). Polymerase chain reactions were carried out under the following conditions: initial denaturation for 3 min at 94 °C; 30 s at 94 °C, 1 min at 60 °C, 3 min at 68 °C as a second step with as many repetitions as required (specified under results), and a final extension for 10 min at 68 °C.

5.6. Analysis of DNA

DNA sequencing was carried out by the didesoxynucleotide chain termination method (Sanger et al., 1977) using an ABI PRISM 377 DNA Sequencer (Applied Biosystems, Rodgau-Jügesheim, Germany, at Department I – Botanik, LMU München). Nucleotide sequences were analyzed and compared with the Sequencher 4.2 program (Gene Codes Cooperation, Ann Arbor, MI, USA).

5.7. Expression in E. coli and isolation of recombinant enzymes

Full-length cDNAs were cloned into a modified pT7-7 vector to give fusion proteins containing N-terminal His_6-tags. The resulting plasmids were transformed into *E. coli* BL21(DE3) cells. Cultures were grown until OD ~ 1 was reached, induced with 0.3 mM isopropyl-β-D-thiogalactopyranoside, and incubated for 5 h at 27 °C. Recombinant enzymes were purified by immobilized metal chelate affinity chromatography using the Ni-NTA metal affinity matrix (Qiagen) according to the instructions of the manufacturer. The affinity matrix was washed with buffer containing 20 mM imidazole followed by elution of adsorbed proteins with buffer containing 250 mM imidazole.

5.8. Enzyme assay

Enzyme activity was determined spectrophotometrically according to the method of Knobloch and Hahlbrock (1975). The change in absorbance caused by CoA-ester formation in the presence of purified recombinant 4CLs was monitored at $\lambda = 311$ nm for cinnamic (**1**) acid, 333 nm for 4-coumaric acid (**2**), 346 nm for caffeic (**3**) and ferulic acids (**4**), and at 352 nm for sinapic acid (**5**) (Gross and Zenk, 1966). The K_m and V_{max} values of recombinant 4CL1, 4CL2, and 4CL3 were determined using Linewaever–Burk plots with at least 10 data points. Relative V_{max} values were obtained by setting V_{max} of 4-coumarate for each isoform to 100%.

5.9. Phylogenetic analysis

4CL amino acid sequences for the phylogenetic reconstruction were retrieved from Genbank (http://www.ncbi.nlm.nih.gov). Gen-

Bank accession nos. for amino acid sequences analyzed in addition to those from *P. patens* were: U18675 (*At*4CL1, *Arabidopsis thaliana*), AF106086 (*At*4CL2, *A. thaliana*), AF106088 (*At*4CL3, *A. thaliana*), AAM19949 (*At*4CL4, *A. thaliana*), AF279267 (*Gm*4CL1, *Glycine max*), AF002259 (*Gm*4CL2, *G. max*), AF002258 (*Gm*4CL3, *G. max*), X69955 (*Gm*4CL4, *G. max*), D49366 (*Le*4CL1, *Lithospermum erythrorhizon*), D49367 (*Le*4CL2, *L. erythrorhizon*), AF052221 (*Lp*4CL1, *Lolium perenne*), AF052222 (*Lp*4CL2, *L. perenne*), AF052223 (*Lp*4CL3, *L. perenne*), D43773 (*Nt*4CL, *Nicotiana tabacum*), U50845 (*Nt*4CL1, *N. tabacum*), U50846 (*Nt*4CL2, *N. tabacum*), X52623 (*Os*4CL1, *Oryza sativa*), L43362 (*Os*4CL2, *O. sativa*), X13324 (*Pc*4CL1, *Petroselinum crispum*), X13325 (*Pc*4CL2, *P. crispum*), U12012 (*Pt*4CL1, *Pinus taeda*), U12013 (*Pt*4CL2, *P. taeda*), AF008184 (*Poph*4CL1, *Populus hybrida*), AF008183 (*Poph*4CL2, *P. hybrida*), AF041049 (*Popt*4CL1, *Populus tremuloides*), AF041050 (*Popt*4CL2, *P. tremuloides*), AF239687 (*Ri*4CL1, *Rubus idaeus*), AF239686 (*Ri*4CL2, *R. idaeus*), AF239685 (*Ri*4CL3, *R. idaeus*), M62755 (*St*4CL1, *Solanum tuberosum*), AF150686 (*St*4CL2, *S. tuberosum*), AL939119 (*Sc*CCL, *Streptomyces coelicolor* A3(2)).

Amino acid sequences were aligned using Clustal W under manual control as implemented in BioEdit (Hall, 1999). The resulting data matrix was subsequently analyzed using PAUP version 4.0 (Swofford, 1998) as reported earlier (Lindermayr et al., 2002). Bayesian analysis (Ronquist and Huelsenbeck, 2003) was performed using two analyses of 10,000,000 generations each and sampling trees every 1000 generations, a "burn-in" of 25%, and equal state frequencies and mutation rates. The 50% majority rule consensus tree was in accordance to the maximum parsimony analysis. The topology is shown in Fig. 4, indicating posterior probability values (Bayes) and bootstrap values (PAUP), respectively. Similarity analysis between *Pp*4CL sequences was performed employing the protein sequence parsimony method protpars using the Phylip program package (Felsenstein, 1989; http://evolution.genetics.washington.edu/phylip.html).

Acknowledgements

The *Physcomitrella* BAC library was constructed at Southern Illinois University (SIU), USA by Khalid Meksem for the Leeds University/Washington University *Physcomitrella* EST program (PEP). We thank PEP and SIU for providing EST and BAC samples. This work was supported by the Deutsche Forschungsgemeinschaft within the priority program 1152 "Evolution of Metabolic Diversity" (Grants EB 62/15-1 and EB 62/15-2).

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Asakawa, Y., 2001. Recent advances in phytochemistry of bryophytes-acetogenins, terpenoids and bis(bibenzyl)s from selected Japanese, Taiwanese, New Zealand, Argentinean and European liverworts. Phytochemistry 56, 297–312.

Ashton, N.W., Grimsley, N.H., Cove, D.J., 1979. Analysis of gametophytic development in the moss, *Physcomitrella patens*, using auxin and cytokinin resistant mutants. Planta 144, 427–435.

Becker-Andre, M., Schulze-Lefert, P., Hahlbrock, K., 1991. Structural comparison, modes of expression, and putative cis-acting elements of the two 4-coumarate:CoA ligase genes in potato. J. Biol. Chem. 266, 8551–8559.

Costa, M.A., Collins, R.E., Anterola, A.M., Cochrane, F.C., Davin, L.B., Lewis, N.G., 2003. An in silico assessment of gene function and organization of the phenylpropanoid pathway metabolic networks in *Arabidopsis thaliana* and limitations thereof. Phytochemistry 64, 1097–1112.

Costa, M.A., Bedgar, D.L., Moinuddin, S.G.A., Kim, K.-W., Cardenas, C.L., Cochrane, F.C., Shockey, J.M., Helms, G.L., Amakura, Y., Takahashi, H., Milhollan, J.K., Davin, L.B., Browse, J., Lewis, N.G., 2005. Characterization *in vitro* and *in vivo* of the putative multigene 4-coumarate:CoA ligase network in *Arabidopsis*: syringyl lignin and sinapate/sinapyl alcohol derivative formation. Phytochemistry 66, 2072–2091.

Doyle, J., Doyle, L., 1990. Isolation of plant DNA from fresh tissue. Focus 12, 13–15.

Ehlting, J., Büttner, D., Wang, Q., Douglas, C.J., Somssich, I.E., Kombrink, E., 1999. Three 4-coumarate:coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionary classes in angiosperms. Plant J. 19, 9–20.

Ehlting, J., Hamberger, B., Million-Rousseau, R., Werck-Reichhart, D., 2006. Cytochromes P450 in phenolic metabolism. Phytochem. Rev. 5, 239–270.

Felsenstein, J., 1989. PHYLIP-Phylogeny Inference Package (Version 3.2). Cladistics 5, 164–166.

Gross, G.G., Zenk, M.H., 1966. Darstellung und Eigenschaften von Coenzym A-Thiolestern substituierter Zimtsäuren. Z. Naturforsch. 21b, 683–690.

Hall, T.A., 1999. BioEdit. A user-friendly biological sequence alignment and analysis program for Windows 95/98/NT. Nucleic Acid Symp. Ser. 41, 95–98.

Hamberger, B., Hahlbrock, K., 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.

Hu, W.-J., Kawaoka, A., Tsai, C.-J., Lung, J., Osakabe, K., Ebinuma, H., Chiang, V.L., 1998. Compartmentalized expression of two structurally and functionally distinct 4-coumarate:CoA ligase genes in aspen (*Populus tremuloides*). Proc. Natl. Acad. Sci. USA 95, 5407–5412.

Jiang, C., Schommer, C.K., Kim, S.Y., Suh, D.-Y., 2006. Cloning and characterization of chalcone synthase from the moss, *Physcomitrella patens*. Phytochemistry 67, 2531–2540.

Jost, W., Baur, A., Nick, P., Reski, R., Gorr, G., 2004. A large plant beta-tubulin family with minimal C-terminal variation but differences in expression. Gene 340, 151–160.

Kaneko, M., Ohnishi, Y., Horinouchi, S., 2003. Cinnamate:coenzyme A ligase from the filamentous bacterium *Streptomyces coelicolor* A3(2). J. Bacteriol. 185, 20–27.

Knobloch, K.H., Hahlbrock, K., 1975. Isoenzymes of *p*-coumarate:CoA ligase from cell suspension cultures of *Glycine max*. Eur. J. Biochem. 52, 311–320.

Koo, A.J.K., Chung, H.S., Kobayashi, Y., Howe, G.A., 2006. Identification of a peroxisomal acyl-activating enzyme involved in the biosynthesis of jasmonic acid in Arabidopsis. J. Biol. Chem. 281, 33511–33520.

Kumar, A., Ellis, B.E., 2003. 4-Coumarate:CoA ligase gene family in *Rubus idaeus*: cDNA structures, evolution, and expression. Plant Mol. Biol. 31, 327–340.

Lang, D., Eisinger, J., Reski, R., Rensing, S., 2005. Representation and high-quality annotation of the *Physcomitrella patens* transcriptome demonstrates a high proportion of proteins involved in metabolism in mosses. Plant Biol. 7, 238–250.

Lee, D., Douglas, C.J., 1996. Two divergent members of a tobacco 4-coumarate:CoA ligase (*4CL*) gene family. Plant Physiol. 112, 193–205.

Lindermayr, C., Möllers, B., Fliegmann, J., Uhlmann, A., Lottspeich, F., Meimberg, H., Ebel, J., 2002. Divergent members of a soybean (*Glycine max* L.) 4-coumarate:coenzyme A ligase gene family. Eur. J. Biochem. 269, 1304–1315.

Lindermayr, C., Fliegmann, J., Ebel, J., 2003. Deletion of a single amino acid residue from different 4-coumarate:CoA ligases from soybean results in the generation of new substrate specificities. J. Biol. Chem. 278, 2781–2786.

Lozoya, E., Hoffmann, H., Douglas, C.J., Schulz, W., Scheel, D., Hahlbrock, K., 1988. Primary structures and catalytic properties of isoenzymes encoded by two 4-coumarate:CoA ligase genes in parsley. Eur. J. Biochem. 176, 661–667.

Markham, K.R., 1988. Distribution of flavonoids in the lower plants and its evolutionary significance. In: Harborne, J.B. (Ed.), The Flavonoids. Chapman and Hall, London, pp. 427–468.

Nishiyama, T., Fujita, T., Shin-I, T., Seki, M., Nishide, H., Uchiyama, I., Kamiya, A., Carninci, P., Hayashizaki, Y., Shinozaki, K., Kohara, Y., Hasebe, M., 2003. Comparative genomics of *Physcomitrella patens* gametophytic transcriptome and *Arabidopsis thaliana*: implication for land plant evolution. Proc. Natl. Acad. Sci. USA 100, 8007–8012.

Ragg, H., Kuhn, D.N., Hahlbrock, K., 1981. Coordinated regulation of 4-coumarate:CoA ligase and phenylalanine ammonia-lyase mRNAs in cultured plant cells. J. Biol. Chem. 256, 10061–10065.

Rensing, S.A., Ick, J., Fawcett, J.A., Lang, D., Zimmer, A., Van de Peer, Y., Reski, R., 2007. An ancient genome duplication contributed to the abundance of metabolic genes in the moss *Physcomitrella patens*. BMC Evol. Biol. 7, 130.

Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.-F., Lindquist, E.A., Kamisugi, Y., Tanahashi, T., Sakakibara, K., Fujita, T., Oishi, K., Shin-I, T., Kuroki, Y., Toyoda, A., Suzuki, Y., Hashimoto, S., Yamaguchi, K., Sugano, S., Kohara, Y., Fujiyama, A., Anterola, A., Aoki, S., Ashton, N., Barbazuk, W.B., Barker, E., Bennetzen, J.L., Blankenship, R., Hyun Cho, S., Dutcher, S.K., Estelle, M., Fawcett, J.A., Gundlach, H., Hanada, K., Heyl, A., Hicks, K.A., Hughes, J., Lohr, M., Mayer, K., Melkozernov, A., Murata, T., Nelson, D.R., Pils, B., Prigge, M., Reiss, B., Renner, T., Rombauts, S., Rushton, P.J., Sanderfoot, A., Schween, G., Shiu, S.-H., Stueber, K., Theodoulou, F.L., Tu, H., Van de Peer, Y., Verrier, P.J., Waters, E., Wood, A., Yang, L., Cove, D., Cuming, A.C., Hasebe, M., Lucas, S., Mishler, B.D., Reski, R., Grigoriev, I.V., Quatrano, R.S., Boore, J.L., 2008. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. Science 319, 64–69.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463–5467.

Schaefer, D.G., Zrÿd, J.-P., 2001. The moss *Physcomitrella patens*, now and then. Plant Physiol. 127, 1430–1438.

Schneider, K., Hövel, K., Witzel, K., Hamberger, B., Schomburg, D., Kombrink, E., Stuible, H.-P., 2003. The substrate specificity-determining amino acid code of 4-coumarate:CoA ligase. Proc. Natl. Acad. Sci. USA 100, 8601–8606.

Schneider, K., Kienow, L., Schmelzer, E., Colby, T., Bartsch, M., Miersch, O., Wasternack, C., Kombrink, E., Stuible, H.-P., 2005. A new type of peroxisomal

acyl-coenzyme A synthetase from *Arabidopsis thaliana* has the catalytic capacity to activate biosynthetic precursors of jasmonic acid. J. Biol. Chem. 280, 13962–13972.

Stuible, H.-P., Kombrink, E., 2001. Identification of the substrate specificity-conferring amino acid residues of 4-coumarate:coenzyme A ligase allows the rational design of mutant enzymes with new catalytic properties. J. Biol. Chem. 276, 26893–26897.

Swofford, D.L., 1998. PAUP: Phylogenetic Analysis Using Parsimony, Version 4.0b1. Computer Program Distributed by the Illinois Natural History Survey, Champaign, IL, USA.

Yazaki, K., Ogawa, A., Tabata, M., 1995. Isolation and characterization of two cDNAs encoding 4-coumarate:CoA ligase in *Lithospermum* cell cultures. Plant Cell Physiol. 36, 1319–1329.