Available online at www.sciencedirect.com
ScienceDirect
Phytochemistry 67 (2006) 2531–2540
PHYTOCHEMISTRY
www.elsevier.com/locate/phytochem

Cloning and characterization of chalcone synthase from the moss, *Physcomitrella patens*

Chenguang Jiang, Clark K. Schommer, Sun Young Kim, Dae-Yeon Suh *

Department of Chemistry and Biochemistry, University of Regina, 3737 Wascana Parkway, Regina, SK, Canada S4S 0A2

Received 30 May 2006; received in revised form 9 September 2006

Abstract

Since the early evolution of land plants from primitive green algae, flavonoids have played an important role as UV protective pigments in plants. Flavonoids occur in liverworts and mosses, and the first committed step in the flavonoid biosynthesis is catalyzed by chalcone synthase (CHS). Although higher plant CHSs have been extensively studied, little information is available on the enzymes from bryophytes. Here we report the cloning and characterization of CHS from the moss, *Physcomitrella patens*. Taking advantage of the available *P. patens* EST sequences, a CHS (PpCHS) was cloned from the gametophores of *P. patens*, and heterologously expressed in *Escherichia coli*. PpCHS exhibited similar kinetic properties and substrate preference profile to those of higher plant CHS. *p*-Coumaroyl-CoA was the most preferred substrate, suggesting that PpCHS is a naringenin chalcone producing CHS. Consistent with the evolutionary position of the moss, phylogenetic analysis placed PpCHS at the base of the plant CHS clade, next to the microorganism CHS-like gene products. Therefore, PpCHS likely represents a modern day version of one of the oldest CHSs that appeared on earth. Further, sequence analysis of the *P. patens* EST and genome databases revealed the presence of a CHS multigene family in the moss as well as the 3′-end heterogeneity of a CHS gene. Of the 19 putative CHS genes, 10 genes are expressed and have corresponding ESTs in the databases. A possibility of the functional divergence of the multiple CHS genes in the moss is discussed.

Keywords: *Physcomitrella patens*; Funariaceae; Moss; Chalcone synthase; Multigene family; Enzyme evolution

1. Introduction

Flavonoids, ubiquitous in higher plants, are also produced by a high proportion of the bryophytes. While no flavonoids have been reported present in hornworts (Anthocerotales), various flavonoids were found in 40% of the liverworts (Hepaticae) and 48% of the mosses (Musci) tested (Markham, 1988). The flavonoid types in mosses include flavone *C*- and *O*-glycosides (Basile et al., 2003), biflavones (Brinkmeier et al., 1999), and aurones (Geiger and Markham, 1992), isoflavones, and 3-deoxyanthocyanins (Markham, 1988). Flavonoids may have played a significant role in the early evolution of land plants, first as chemical messengers and then as UV filters (Stafford, 1991; Winkel-Shirley, 2001). Evolutionary studies support the monophyletic origin of land plants. Thus, bryophytes, which diverged from the ancestors of vascular plants ~450 million year ago, are ideal for comparative studies of the evolution of biosynthesis pathways and other biological processes in land plants. However, only a limited information is available on the enzymes of flavonoid biosynthesis in the bryophytes.

Chalcone synthase (CHS, EC 2.3.1.74) catalyzes the first committed step of flavonoid biosynthesis. It first condenses a phenylpropanoid CoA-ester (e.g. *p*-coumaroyl-CoA (**1**)) with three C_2 units from malonyl-CoA, and cyclizes the resulting tetraketide intermediate to give a chalcone (e.g. naringenin chalcone (**2**)) (Fig. 1) (Schröder, 1997). Several hundred CHS genes have been cloned from plants, and the structure and reaction mechanism of higher plant CHSs have been extensively studied (Ferrer et al., 1999; Suh et al.,

* Corresponding author. Tel.: +1 306 585 4239; fax: +1 306 337 2409.
E-mail address: suhdaey@uregina.ca (D.-Y. Suh).

0031-9422/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2006.09.030

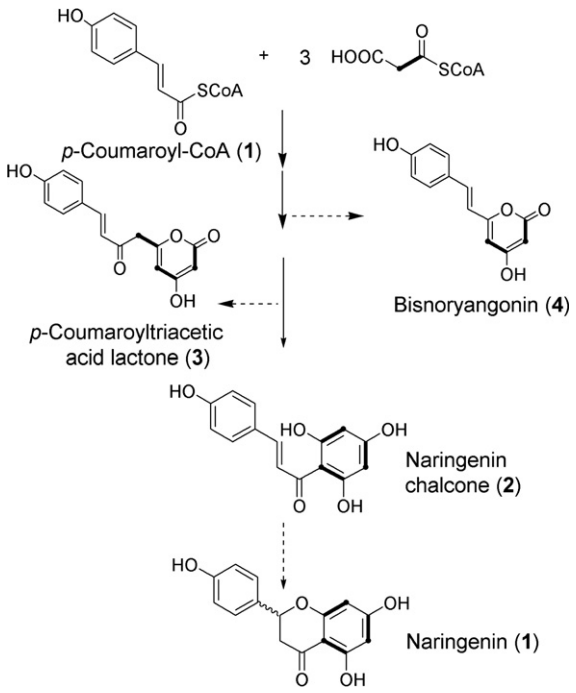

Fig. 1. Reaction scheme of chalcone synthase (CHS). Bisnoryangonin (**4**) and *p*-coumaroyltriacetic acid lactone (**3**) are *in vitro* derailment products released from the enzyme active site after two and three condensations, respectively. The CHS product naringenin chalcone (**2**) is spontaneously converted to naringenin (**5**) under the reaction conditions used. Chemical bonds originating from the C_2 units of malonyl-CoA are indicated by thick lines.

2000; Jez and Noel, 2000). However, the enzyme properties and the mechanism of gene regulation of CHS from bryophytes are scarcely studied, although the CHS activity has been observed in crude extracts of the liverwort, *Marchantia polymorpha* (Fischer et al., 1995). Recently a CHS-like gene was isolated from the liverwort, *Marchantia paleacea*, and its gene product (presumably a CHS) was shown to be expressed light-dependently in photoautotrophic cells (Harashima et al., 2004).

The moss *Physcomitrella patens* (Funariaceae) has drawn much attention because of its evolutionary position and exceptional genetic properties including accessibility to gene targeting (Schaefer and Zrÿd, 2001). Thus, thanks to multiple sequencing projects, nucleic acid sequences of >100,000 ESTs (expressed sequence tag) and >16,000 annotated transcripts are available (Nishiyama et al., 2003; Lang et al., 2005). Moreover, the first draft sequence of the *P. patens* genome has recently been released and is in the process of gene annotation at the time of writing. The genomic sequence (*Physcomitrella patens v.1.0*) assembled from ~5.5 million shotgun reads is composed of a total of 2536 scaffolds containing 21,949 contigs (http://shake.jgi-psf.org/Phypa1/Phypa1.home.html). In this study, a CHS gene was cloned from the gametophore tissues of *P. patens*. The moss CHS was overexpressed in *Escherichia coli*, and its kinetic properties and substrate preference were compared to those of a higher plant CHS. In addition, all available *P. patens* EST and genome sequences were analyzed to reveal the presence of multiple CHS superfamily genes in the moss.

2. Results

2.1. cDNA isolation and sequence analysis

From the sequences of *P. patens* ESTs that were annotated as putative CHS, a set of gene specific primers were designed. PCR was performed with these primers and cDNA isolated from three week-old gametophore tissues to amplify a 451 bp core fragment. The full length cDNA including the 5′- and 3′-untranslated regions (UTR) was obtained by 5′- and 3′-RACE (accession no. DQ275627). The cDNA contained an open reading frame (ORF) of 1194 bp encoding a 42.6 kDa protein of 397 amino acids (hereafter referred as PpCHS, accession no. ABB84257). The nucleotide sequence surrounding the designated start codon (TCCAATGGC) is similar to the initiation consensus sequence of AACAATGGC in plants (Lütcke et al., 1987). The presence of an in-frame stop codon (TAG) 30 bp upstream of the start codon and the absence of any other start codon in the 5′-UTR (278 bp) lend further supports that the designated ATG is the real start codon. The ATG start codon is followed by GCTTCT, thus the second and the third residues are Ala and Ser, which are the predominant residues at corresponding positions in highly expressed plant genes. Sawant et al. (2001) found the codon GCT immediate downstream of the start codon in 95% of proteins including photosystem related proteins and Phe ammonia lyases that were classified as highly expressed plant proteins in the literature.

The strictly conserved CHS active site residues, Cys^{170}, His^{309}, and Asn^{342} (Ferrer et al., 1999) as well as the highly conserved CHS signature sequence, G^{378}FGPG, (Suh et al., 2000) were found in PpCHS (Fig. 2). The two Phe residues (Phe^{221} and Phe^{271}), important in determining the substrate specificity of CHS (Jez et al., 2002), were also found. The PpCHS amino acid sequence showed the highest sequence identity to a liverwort CHS-like gene product (BAD42329) (67% identity, 88% similarity), reflecting their close evolutionary relationship. The sequence identity was determined to be 62% as compared to other primitive plant CHSs, *Psilotum nudum* CHS (BAA87922) and *Equisetum arvense* CHS (BAA89501), and generally in the range of 60 ~ 65% to higher plant CHSs.

2.2. Functional expression of PpCHS in E. coli, purification, and enzyme assay

To confirm the functional identity of PpCHS, the corresponding ORF was cloned into the *E. coli* expression vector pET-32a(+). Thus, PpCHS was overproduced as a thioredoxin (Trx)-fusion protein (~62 kDa) and the $(His)_6$-tag enabled us to obtain the enzyme of high purity after a single purification step of Ni^{2+}-chelation chroma-

```
Arabidopsis thaliana   -MMYQQGCFAGG- -FWIAHPGGPA- -SEYGNMSSAC- -LFGFGPGLT-
Alfalfa (M. sativa)    -MMYQQGCFAGG- -FWIAHPGGPA- -SEYGNMSSAC- -LFGFGPGLT-
Pine (P. sylvestris)   -MMYQQGCFAGG- -FWIAHPGGPA- -SDYGNMSSAC- -LFGFGPGLT-
Horsetail (E. arvense) -MMYQQGCFAGG- -FWIAHPGGPA- -SEYGNMSSAC- -LLGFGPGLT-
Whisk fern (P. nudum)  -MLYQQGCFGGG- -FWIAHPGGPA- -ADYGNMSSAC- -LFGFGPGLT-
Livewort (M. paleacea) -MLYQQGCFGGA- -FWCVHPGGRA- -YNYGNMSGAS- -VVGFGPGLT-
Moss (P. patense)      -MMYQTGCFGGA- -FWAVHPGGPA- -SEFGNMSSAS- -FIGFGPGLT-
                             170           309           342           381
```

Fig. 2. Alignment of the parts of the CHS sequences from *Arabidopsis thaliana* (CAI30418), *Medicago sativa* (P30074), *Pinus sylvestris* (CAA43166), *Equisetum arvense* (Q9MBB1), *Psilotum nudum* (BAA87922), *Marchantia paleacea* (BAD42328), and *Physcomitrella patens* (ABB84527). Sequences around the active site residues, Cys, His, and Asn, and the signature GFGPG loop (Suh et al., 2000) are shown. The residue numbers are those of *P. patens* chalcone synthase.

tography (Fig. 3A). Under the conditions employed, a good portion of the overproduced PpCHS in *E. coli* was recovered in soluble fractions with an overall yield of 4 mg/100 ml culture. The Trx and $(His)_6$-tag are joined with the CHS polypeptide at its N-end via a linker containing an enterokinase cleavage site. Earlier studies showed that there is no functional difference between the Trx-fusion enzyme and the native enzyme recovered by cleavage of the Trx-fusion enzyme with enterokinase (Suh et al., 2000). Thus the thioredoxin fusion protein was used for further analysis.

In vitro enzyme assay using *p*-coumaroyl-CoA (**1**) and [2-^{14}C] malonyl-CoA as substrates confirmed that PpCHS is indeed a CHS. PpCHS produced naringenin chalcone (**2**) as a major product as well as the derailment products, *p*-coumaroyltriacetic acid lactone (**3**, CTAL, the derailed lactone after three condensations) and bisnoryangonin (**4**, BNY, the derailed lactone after two condensations) (Figs. 1 and 3B). The reaction products were identified by either using internal standard or by their known R_f values

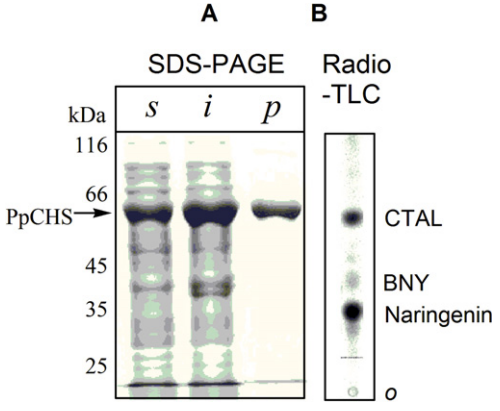

Fig. 3. Overproduction, purification, and enzyme assay of *P. patens* CHS (PpCHS). (A) Soluble (*s*) and insoluble (*i*) *E. coli* extracts and purified PpCHS (*p*) were analyzed by SDS-polyacrylamide gel electrophoresis. The proteins were separated on 12% gel under reducing conditions, and the positions of molecular mass markers are indicated on the left. PpCHS was expressed as a thioredoxin fusion protein (62 kDa). (B) Radio thin-layer chromatogram of reaction products produced by PpCHS. Enzyme activity of PpCHS was determined with *p*-coumarol-CoA (**1**) and [2-^{14}C]malonyl-CoA as substrates as described in *Experimental* section. Derailment products, BNY (**4**) and CTAL (**3**), were produced as well as naringenin chalcone (**2**), the cyclization product of CHS, which was spontaneously converted to naringenin (**5**) during the reaction work-up.

(Yamaguchi et al., 1999). Enzymatic properties of PpCHS were compared to those of CHS from a higher plant, *Pueraria lobata*. PpCHS was indistinguishable to *P. lobata* CHS in product profiles (Table 1) and catalytic competency. The specific activity of PpCHS was 50 pkat/mg, comparable to the activity of *P. lobata* CHS (57 pkat/mg).

2.3. Steady-state kinetic analysis and starter-CoA preference of PpCHS

PpCHS was further characterized for its kinetic properties. The $K_{m(app)}$ values for *p*-coumaroyl-CoA (**1**) and malonyl-CoA of PpCHS were 48 ± 5.2 μM ($n = 3$, mean $\pm$ standard deviation) and 3.5 ± 0.25 μM, respectively (Table 1). Again, these values were comparable to those of *P. lobata* CHS obtained under identical conditions (Suh et al., 2000). Then, the starter-CoA preference profiles of PpCHS and *P. lobata* CHS were compared using various CoA esters as the starter substrate. The cyclization products (Fig. 4, marked by asterisks) were identified based on their independence of the pH during ethyl acetate extraction and on the comparison with the published data (Yamazaki et al., 2001).

Among the CoA esters examined, *p*-coumaroyl-CoA (**1**) was by far the most preferred substrate for the cyclization reaction of PpCHS (Table 2). No study has been reported on the types of flavonoids produced in *P. patens*. In our preliminary study, naringenin (**5**) was detected both in alcoholic extracts (EtOH-H_2O (7:3, v/v), overnight at 25 °C) and in acid extracts (boiled in 2 N HCl for 1 h) of 3–6 week old *P. patens* gametophores. Naringenin (**5**) in the extracts was identified by HPLC. Both alcoholic and acid extracts produced a well separated 27.5 min peak that co-eluted with authentic naringenin (**5**) in RP-HPLC analysis (Apollo C18 5 μ, 4.6 × 250 mm; flow rate, 1.0 ml/min; UV 290 nm; CH_3CN in H_2O, 30 to 45% for 10 min, 45 to 60% for another 30 min). Together with the finding that cinnamoyl-CoA and dihydro-*p*-coumaroyl-CoA were converted to the corresponding chalcones by PpCHS at 50 ~ 60% efficiency as compared to *p*-coumaroyl-CoA (**1**), these results suggested that *p*-coumaroyl-CoA (**1**) is likely the *in vivo* substrate for PpCHS. Both PpCHS and *P. lobata* CHS exhibited a general preference towards phenylpropanoid-CoA esters, a hallmark characteristic for plant CHS (Table 2).

Table 1
Comparison of product profiles and kinetic parameters of chalcone synthases from the moss, *P. patens* and a higher plant, *Pueraria lobata*

	Products produced (mol%)			Kinetic parameters (mean ± SD)		
				$K_{m(app)}$ (μM)		V_{max} (pmol s^{-1} mg^{-1})
	Chalcone (**2**)	BNY (**4**)	CTAL (**3**)	*p*-Coumaroyl-CoA (**1**)	Malonyl-CoA	
PpCHS	70	7	23	48 ± 5.2	3.5 ± 0.25	0.68 ± 0.085
P. lobata CHS	66	8	26	50 ± 7.1[a]	3.0 ± 1.2[a]	0.46 ± 0.27[a]

[a] From the reference, Suh et al. (2000).

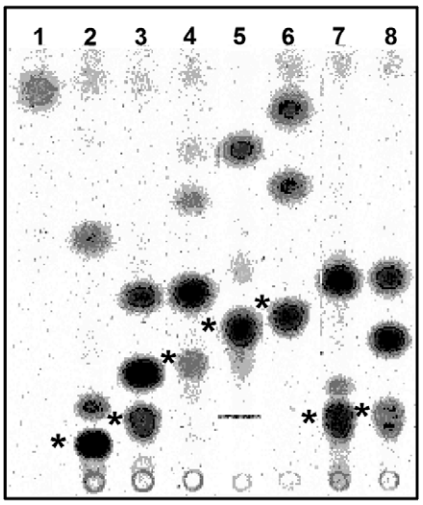

Fig. 4. Radio thin-layer chromatogram of reaction products produced by PpCHS. Starter CoA used was: Lane 1, acetyl-CoA; 2, butyryl-CoA; 3, hexanoyl-CoA; 4, isovaleryl-CoA; 5, *p*-coumaroyl-CoA (**1**); 6, dihydro-*p*-coumaroyl-CoA; 7, cinnamoyl-CoA, and 8, dihydro-cinnamoyl-CoA. The enzyme reaction was run with 100 μM starter CoA and 17 μM [2-^{14}C]malonyl-CoA, and the reaction products were analyzed by RP-TLC as described in *Experimental*. Cyclization products, marked with asterisks (*), were assigned on the basis of their behaviour during extraction at different pH values as described in the text.

2.4. Putative CHS superfamily genes from P. patens

In the *P. patens* EST databases, we found a total of 41 EST sequences with significant clarity and length that were annotated as putative CHS superfamily genes; 8 in GenBank, 17 in the Physcomitrella EST Programme (Lang et al., 2005), and another 16 in the PHYSCObase (Nishiyama et al., 2003). Multiple EST sequences were derived from identical transcript; however, the presence of multiple CHS genes in *P. patens* was evident. In the course of this study, the first draft sequence of the *P. patens* genome became available. Extensive search led us to find 19 putative CHS superfamily genes in *P. patens* (Table 3). The sequences of four of these genes were only partially available, but their sequence similarity still allowed us to assign them as putative CHS genes. These genes were named from *PpCHS1a* to *PpCHS14* in such a way that *PpCHS1a* and *PpCHS1b* share the same coding sequence but different UTR sequences, and that *PpCHS5.1* and *PpCHS5.2* are two copies of the same gene found in different locations. These 19 genes accounted for all the 41 ESTs except one EST, BJ610253, whose sequence is similar, but not identical to *PpCHS4*. Ten of the 19 genes are expressed, as they were represented by ESTs (Table 3). PpCHS cloned in this study is encoded by *PpCHS13a*. Of the other 18 genes, 13 genes belonging to 8 gene classes (*PpCHS1* to *7*, and *13*) showed amino acid sequence identity of >85% as compared to PpCHS, suggesting that they may also encode CHS. However, the possibility that some of these genes encode enzymes that accept different CoA substrates or catalyze different numbers of condensation reaction cannot be excluded. Meanwhile, the sequence identity of the remaining 5 genes varied from 52.6% to 23.1%, suggesting that they may encode other CHS-like enzymes. The 3′-UTR sequences of *PpCHS1a*, 1*b*, 1*d*, and 13*a* were extremely similar, indicating that these genes are the products of recent gene duplication events (see Supplementary data for sequences).

In addition, at least three pseudogenes were also found. Here, we regarded a gene as a pseudogene if there is at least one stop codon or nucleotide deletion in the coding sequence. Their KEGG protein ID and locations are ID 63613, 510:7601–8787; ID 155379, 463:76491–77898, and ID 68833, 20:2010310–2009123. Gene fragments of various lengths are also found at at least 14 different locations (data not shown). For example, a gene found at Scaffold 38 (399922–41402) encodes a 271 amino acid long polypeptide that contains the catalytic Cys, His and Asn triad, but not the GFGPG sequence. A gene fragment found at Scaffold 232 (292297–388929) contains parts of the 5′-UTR and coding sequences of the EST, BJ610253. However, we failed to find the full sequence of BJ610253 in the genome.

Table 2
Starter CoA preference of chalcone synthases from the moss, *P. patens* and a higher plant, *P. lobata*

Starter CoA	Relative activity[a] (%)	
	PpCHS	*P. lobata* CHS
p-Coumaroyl-CoA (**1**)	(100)	(100)
Acetyl-CoA	0	0
Butyryl-CoA	43	39
Isovaleryl-CoA	15	20
Hexanoyl-CoA	37	19
Dihydro-*p*-couma royl-CoA	50	130
Cinnamoyl-CoA	62	91
Dihydrocinnamoyl-CoA	23	25

[a] Based on the production of cyclization products.

Table 3
Analysis of putative chalcone synthase superfamily genes found in the *P. patens* genome (*version 1.0*) (http://shake.jgi-psf.org/Phypa1/Phypa1.home.html)

Gene	KEGG Protein ID	Location		Sequence identity to ABB84527 (%)	Corresponding EST	Description
		Scaffold	Base position			
A. Putative CHS superfamily genes of which full sequences are available[a]						
PpCHS1a		500	76526–77719	99.2	Contig4041, Contig4179, PPP_7231_C2, BJ177589	No intron.
PpCHS1b		22	12548–13741	99.2	BJ589895, Contig14228, BJ604022	Differ from *PpCHS1a* in 3′-UTR.
PpCHS1c		426	238637–239830	99.2		Differ from *PpCHS1a* in 3′-UTR.
PpCHS2a	101257	463	99838–101304	94	Contig7184, PPP_7231_C1, PPP_1175_C1, PPP_7571_C1	One intron splits Cys^{72}.
PpCHS2b.1	101260	463	129624-131090	94		Differ from *PpCHS2a* in 5′-UTR.
PpCHS2c		1	81194–82600	94	Contig7183, PPP_2112_C1	Differ from *PpCHS2a* in 3′-UTR.
PpCHS3.1	149692	303	69378–70547	92.2		No intron.
PpCHS3.2	149682	303	79392–80561	92.2		Another copy of *PpCHS3.1*.
PpCHS3.3	149180	292	1050–2219	92.2		Another copy of *PpCHS3.1*.
PpCHS4		39	1774478–1775529	91.6		No intron. Similar to EST, BJ610253.
PpCHS5.1	152430	365	131783–132952	90.1	Contig10984, Contig11523,	No intron.
PpCHS5.2	98737	365	151849–153018	90.1	Contig11524, PPP_2843_C1, BJ157102, Contig10983	Another copy of *PpCHS5.1*.
PpCHS6	3283	228	276349–279881	85.5	PPP_7327_C1, Contig467, Contig466, PPP_142_C1	One intron splits Cys^{76}.
PpCHS7	109184	425	82874–84715	85.4		One intron splits Ser^{108} (?). Similar UTR's with *PpCHS1a*.
PpCHS8		25	2487592–2488635	52.6		No intron.
PpCHS9	118540	25	2490141–2491495	42.1		One intron splits Cys^{77}.
PpCHS10	149790	304	419812–421296	38.2		Two introns slpit Cys^{88} and Gly^{333}.
PpCHS11	56368	34	1066367–1068089	33	Contig1663, PPP_6702_C1	Two introns slpit Cys^{80} and Arg^{348}.
PpCHS12	157037	639	40768–41772	23.1		No intron. 5′-UTR sequence unavailable.
B. Putative CHS superfamily genes of which full length sequences are not available						
PpCHS13a	410	292865–293711		100[b]	Contig4180, PPP_182_C3, Contig7312, BJ583475 Contig9794,[c] PPP_182_C2[c]	This study. 49 bp 5′-UTR and 847 bp cds.
PpCHS13b	16	2859768–2860642		100[b]		506 bp 5′-UTR and 369 bp cds. 5′-UTR identical to *PpCHS1a*.
PpCHS1d	16	2867184–2868325		–	Contig4042	Identical to the 3′-UTR of Contig4042.
PpCHS2b.2	76	1642316–1643010		94.2[b]		Another copy of *PpCHS2b.1* (?)

[a] Amino acid sequences of different genes are shown in Supplementary data.
[b] Calculated with the matching sequences only.
[c] Shows a slightly different 3′-UTR sequence that contains nucleotide insertions (see Supplementary data).

All the plant CHS superfamily genes including the CHS-like gene from the liverwort, Marchantia, contain an intron at a conserved Cys (see Fig. 3B in Harashima et al., 2004). Surprisingly, the *P. patens* CHS superfamily genes showed a complex pattern with regard to intron (Table 3). While all but one of the genes with one intron have the intron/exon site at the conserved Cys, two genes (*PpCHS10* and 11) have an additional intron/exon site close to their C-ends. To our knowledge, this is the first report of the CHS superfamily genes with either no or two introns.

In plants, the position of cleavage and polyadenylation of mRNA 3′-end can be heterogeneous within a single transcript (Rothnie, 1996). The hexamer motif AAUAAA, known to be required for the 3′-end processing in mammalian cells, is less conserved or absent in plants, leading to the 3′-end heterogeneity (Loke et al., 2005). Indeed, our nested 3′-RACE reaction yielded PCR products of various sizes. Of the 9 clones sequenced, four had the same 3′-UTR sequence (Type 3B in Supplementary data), while three other 3′-UTR variants were of various lengths (data not shown). The 3′-termini of all the clones analyzed were followed by the poly(A) tails, suggesting that the observed multiple polyadenylation sites in the gene, *PpCHS13a*, may not be cloning artifacts. Heterogeneous polyadenylation has been observed with higher plant genes such as chloroplast RNA-binding proteins in *Nicotiana plumbaginifolia* (Klahre et al., 1995) and a H^+-ATPase subunit A gene in *Arabidopsis thaliana* (Magnotta and Gogarten, 2002).

2.5. Phylogenetic analysis

In addition to the Bayesian estimation method, we conducted neighbor-joining (NJ) and maximum parsimony (MP) analyses using the ClustalX and MEGA (v. 3.1) software packages (Thompson et al., 1997; Kumar et al., 2004). In all analyses, the position of the moss CHS superfamily in the phylogenetic tree was found to be consistent with

its evolutionary position. Thus, the moss enzymes were located along with the two liverwort CHS superfamily enzymes at the base of the plant CHS clade, in between the CHS-like gene products from microorganisms and the pteridophyte CHS superfamily enzymes (Fig. 5).

CHS is the representative member of the CHS superfamily, also known as the type III polyketide synthases (Shen and Hutchinson, 1993; Austin and Noel, 2003). Once believed to be specific to the plant kingdom, the CHS superfamily now contains growing number of CHS-like enzymes found in microorganisms. *Streptomyces griseus* RppA catalyzes the formation of 1,3,6,8-tetrahydroxynaphthalene from five molecules of malonyl-CoA, while *Psuedomonas fluorescens* PhlD produces monoacetylphloroglucinol from acetoacetyl-CoA and malonyl-CoA (Funa et al., 1999; Bangera and Thomashow, 1999). Enzymes of the CHS superfamily are also involved in the biosynthesis of cell wall lipids in *Mycobacterium tuberculosis* (Saxena et al., 2003) and in the production of a glycopeptide antibiotic balhimycin in *Amycolatopsis mediterranei* (Pfeifer et al., 2001). Gross et al. (2006) conducted an extensive phylogenetic analysis of more than 50 bacterial and fungal CHS-like gene products, and suggested that the plant CHS superfamily enzymes might have been evolved from bacterial CHS-like enzymes. The results obtained from our phylogenetic analysis are in agreement with the proposition of the ancestry of

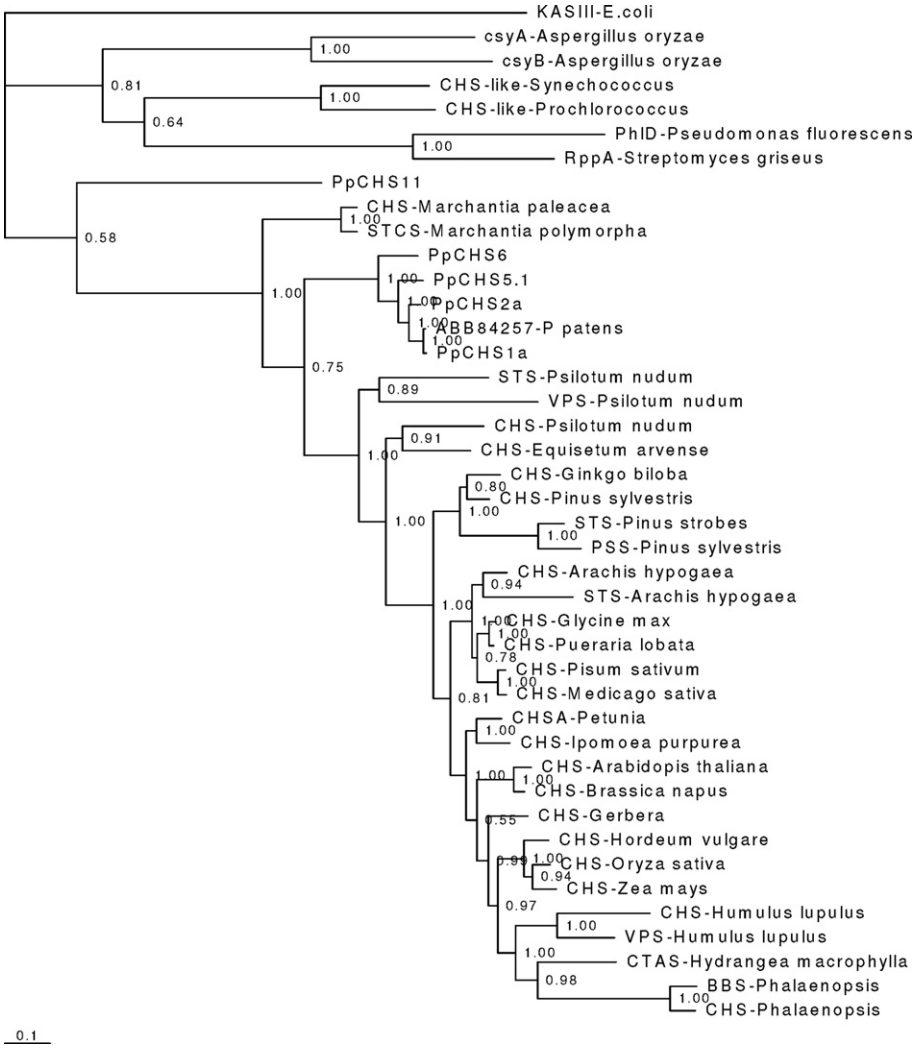

Fig. 5. Amino acid sequence based phylogeny of selected CHS superfamily enzymes from microorganisms and plants, inferred by Bayesian analysis. Numbers above branches indicate the probability of each clade in the tree. Branch length indicates number of expected substitutions per site. *E. coli* β-ketoacyl-(acyl carrier protein) synthase III (KAS III, BAA35899) served as an outgroup. Accession numbers of the enzymes included in the analysis are: *Pseudomonas fluorescens* (AAB48106), *Streptomyces griseus* (BAE07216), *Synechococcus* sp. (CAE07508), *Prochlorococcus marinus* (CAE20587), *Aspergillus oryzae* (csyA, BAD97390; csyB, BAD97391), *Marchantia paleacea* (BAD42328), *Marchantia polymorpha* (STCS, AAW30009), *Physcomitrella patens* (ABB84527), *Psilotum nudum* (CHS, BAA87922; STS, BAA87924; VPS, BAA87923), *Equisetum arvense* (CHS, Q9MBB1), *Pinus strobes* (STS, CAA87013), *Pinus sylvestris* (CHS, CAA43166; PSS, CAA43165), *Ginkgo biloba* (AAT68477), *Arachis hypogaea* (CHS, AAO32821; STS, BAA78617), *Medicago sativa* (P30074), *Pisum sativum* (BAA22044), *Pueraria lobata* (BAA01075), *Glycine max* (AAO67373), *Petunia* × hybrida (P08894), *Ipomoea purpurea* (AAK39115), *Brassica napus* (AAC31913), *Arabidopsis thaliana* (CAI30418), *Gerbera* (CAA86218), *Hordeum vulgare* (CAA41250), *Zea mays* (CAA42763), *Oryza sativa* (BAA19186), *Humulus lupulus* (CHS, BAB47195; VPS, BAB12102), *Hydrangea macrophylla* (CTAS, BAA32733), *Phalaenopsis* sp. (CHS, AAV70116; BBS, CAA56277). STCS, stilbenecarboxylate synthase; STS, stilbene synthase; VPS, valerophenone synthase; PSS, pinosylvin synthase; CTAS, coumaroyltriacetic acid synthase; BBS, bibenzyl synthase.

bacterial CHS-like enzymes for the plant CHS superfamily. In this regard, the gene *PpCHS11* may prove to be informative. *PpCHS11* appeared as the most recent common ancestor to all the plant CHS superfamily genes, although the probability of that particular branch in the phylogenetic tree was relatively low (0.58) (Fig. 5).

3. Discussion

As simple flavonoids (flavanones, flavones and flavanols) are found in bryophytes, it is thought that CHS first appeared in bryophytes during the early evolution of land plants. Therefore, the moss CHS reported in this paper (PpCHS), which shows similar enzymatic characteristics as compared to higher plant CHSs, likely represents a modern day version of one of the oldest CHSs. Together with the liverwort CHS-like genes (Harashima et al., 2004), the moss CHS superfamily provides the long-missing bridge in the phylogenetic tree between CHS-like enzymes from microorganisms and pteridophyte CHS superfamily enzymes. In order to gain further information on the molecular evolutionary process of the CHS superfamily, more studies are needed on the enzymes of the CHS superfamily from primitive plants and from microorganisms, in particular, cyanobacteria and fungi (Seshime et al., 2005).

Amino acid sequence identity between CHS and other CHS superfamily enzymes from *Psilotum nudum* is lower than 60% (Yamazaki et al., 2001), whereas nucleotide sequence identity varies from 78% to 98% among the eight *Pisum sativum* CHS genes (Ito et al., 1997). In the *P. patens* genome, seven closely related gene classes (*PpCHS1* to *PpCHS7*) were found in addition to *PpCHS13* (Table 3). The high degree of sequence identity (>85%) among these genes strongly suggested that these genes likely encode CHS enzymes. Thus, there may be as many as eight CHS gene classes in *P. patens* (*PpCHS1* ~ *7*, *13*) that are expressed at different developmental stages or under different growth conditions. Indeed, five of these gene classes (*PpCHS1*, *2*, *5*, *6*, and *13*) are expressed under the conditions used to collect ESTs. That CHS is encoded by multigene family in higher plants is well documented. A total of 9 CHS genes were found in *Glycine max* (Matsumura et al., 2005), and 7 in *P. sativum* (Ito et al., 1997). The Institute for Genomic Research (TIGR) Rice Database (www.tigr.org) lists 33 putative CHS superfamily genes. The CHS multigene family in higher plants is directly related to the functional diversity of flavonoids. Some of these genes are constitutively expressed while others are transcriptionally induced by environmental factors such as UV light and pathogen attack. As mosses typically exhibit a simpler flavonoid pattern (Stafford, 1991), the physiological significance of the observed CHS multigene family in moss requires further investigation. Of particular interest are the mechanism by which the multiple CHS genes are differentially regulated and the evolutionary relationship and possible functional divergence of these genes. Since *P. patens* is applicable to genetic manipulations including gene targeting, the moss CHS genes may be found useful in studying, for example, light response gene regulation in primitive land plants.

4. Concluding remarks

A naringenin chalcone producing CHS was cloned from the moss, *Physcomitrella patens*, and its enzymatic properties were determined to be very similar to those of a higher plant CHS. Analysis of the *P. patens* EST and genome sequences revealed the presence of a CHS multigene family in the moss as well as the 3′-end heterogeneity of one of the CHS genes (*CHS13a*). Further studies are warranted to investigate the functional divergence and regulation of the CHS multigene family in the moss.

5. Experimental

5.1. Plant material

Gametophores of *Physcomitrella patens* (Hedw.) pabB4 strain were kindly provided by Dr. N. Ashton (University of Regina). Three-week-old tissues, grown on agar plates under continuous light at 16 °C as described previously (Ashton et al., 1985), were collected free of agar, chilled in liquid nitrogen, and stored at −80 °C.

5.2. Cloning of the core fragment of the moss chalcone synthase

Total mRNA was obtained from *P. patens* gametophore tissues (2 g) using the Straight A's mRNA Isolation System (Novagen) as directed by the manufacturer. Reverse transcription-polymerase chain reaction (RT-PCR) was carried out using the SuperScript III One-Step RT-PCR System (Invitrogen) with gene-specific primers of 5′-CCAAGCCTGAGGTAGAGAAGC-3′ and 5′-GTCAGTCCTGGTCCGAAGCC-3′, which were designed based on the *P. patens* EST sequences (accession no. BG409216 and AW509746). The PCR program was: 1 cycle of 30 min at 50 °C and 2 min at 94 °C for cDNA synthesis and pre-denaturation, 40 cycles of 15 s at 94 °C, 30 s at 58 °C, and 1 min at 68 °C for amplification, and 1 cycle of 5 min at 68 °C for final extension. The resulting 451 bp core fragment was cloned into the pCR4-TOPO vector using the TOPO TA Cloning Kit (Invitrogen) for sequencing.

5.3. 5′- and 3′-ends amplification of the P. patens CHS cDNA

Both 5′- and 3′-RACE were performed using the SMART™ RACE cDNA Amplification Kit (Clontech) according to the supplier's instructions. For each RACE experiment, two gene specific primers were synthesized

from the core sequence to allow nested PCR. For 5′-RACE, external primer was 5′-GGTCTCTCCAGCCCAGTGCACTTCGAACAA-3′ and internal primer was 5′-GGCTTCTCTACCTCAGGCTTGGGG-3′. For 3′-RACE, external primer was 5′-CGGGCTTGTGGGCTCAGCTCTGTT-3′ and internal primer was 5′-AAGCCAGTGGGATCACCGGCA-3′. Each PCR product was cloned into pCR4-TOPO vector and sequenced.

5.4. Cloning of the P. patens CHS gene, expression in E. coli, and purification

The ORF of PpCHS was PCR-amplified with the 5′-RACE cDNA mixture as template. The 5′-flanking primer was 5′-TTGTCCCATGGCTTCTGCTGGGGATG (*Nco* I site underlined), and the 3′-flanking primer was 5′-TGGAATTCTAAGCGGAGTTGGGAGCGGCGC (*EcoR* I site underlined). The PCR program was: 3 min of denaturation at 94 °C, 35 cycles of 30 s at 94 °C, 45 s at 64 °C, 45 s at 72 °C, and 7 min of final extension at 72 °C. Thus amplified PCR product was gel-purified, digested with restriction enzymes, and subcloned into the *Nco* I/*EcoR* I-treated pET-32a(+) (Novagen). The resulting plasmid was transformed into the *E. coli* AD494(DE3)-pLysS cells (Novagen) by the heat-shock method. PpCHS was expressed as a thioredoxin-HisTag-fusion protein and purified in 0.1 M potassium phosphate buffer (pH 7.2), 0.1% Triton X-100, 10% glycerol, 1 mM DTT (1,4-dithiothreitol), as described previously (Yamazaki et al., 2001).

5.5. Enzyme assays, steady-state kinetic analysis and starter-CoA preference

For CHS assay, the standard assay mixture (0.1 ml) contained appropriate amounts of purified enzyme ($2 \sim 4$ μg), 0.1 mM starter CoA ester (e.g. *p*-coumaroyl-CoA (**1**)) and 17 μM [2-^{14}C]malonyl-CoA (1.74 GBq/mmol, NEN) in 100 mM potassium phosphate buffer (pH 7.2) containing 0.1% Triton X-100 (enzyme assay buffer). After incubation at 37 °C for 30 min, the reaction was stopped by acidification with 1 N HCl (7.5 μl) and the reaction products were extracted with EtOAc (200 μl). After a brief centrifugation, a portion (50 μl) of the extract was applied on an RP18 plate (Merck 1.15389), and TLC was performed with MeOH:H_2O:AcOH (60:40:1, vol/vol/vol). The radioactive products were quantified with an imaging plate analyzer (Molecular Dynamics Storm 860, Amersham Pharmacia) using standards of known specific activity. Authentic naringenin (**5**, Sigma, $R_f = 0.3$) was used as internal standard to identify the enzyme reaction product. Derailment products, BNY (**4**) and CTAL (**3**), were identified from their published R_f values (BNY, 0.4; CTAL, 0.6) (Yamaguchi et al., 1999). Protein concentration was determined using the protein assay dye reagent (Bio-Rad) with BSA as standard. The specific enzyme activity was expressed in pmol of the product produced s^{-1} mg^{-1} (pkat/mg).

All kinetic experiments were conducted in the enzyme assay buffer at pH 7.2. The K_m values for *p*-coumaroyl-CoA (**1**) were determined at 17 μM malonyl-CoA concentration, and those for malonyl-CoA were determined at 150 μM *p*-coumaroyl-CoA (**1**). At least five different substrate concentrations covering the range of 0.2–4 K_m were used. Obtained data of v against [S] were fitted to the Michaelis–Menten equation using a non-linear regression program (Enzyme Kinetics Pro, ChemSW) to calculate V_{max} and K_m values.

5.6. Preparation of starter CoA esters

p-Coumaroyl-CoA (**1**) and other starter CoA esters were synthesized from the respective *N*-hydroxysuccinimide esters according to the active ester exchange method by Stöckigt and Zenk (1975). The CoA esters were purified with RP-HPLC (Apollo C18 5 μ, 4.6 × 250 mm; flow rate, 0.6 ml/min; UV 260 nm). The chromatography was run using a gradient of MeOH (40–80% in 30 min) in water. Under these conditions, the retention times of *p*-coumaroyl-CoA (**1**), dihydro-*p*-coumaroyl-CoA, cinnamoyl-CoA, dihydrocinnamoyl-CoA were 6.0, 5.7, 7.8, and 6.8 min, respectively.

5.7. Analysis of the P. patens CHS superfamily genes

To find EST sequences of putative CHS superfamily genes from *P. patens*, a BLAST search was performed with both the UTR and coding sequences of PpCHS against the GenBank, the Physcomitrella EST Programme (www.cosmoss.org), and the PHYSCObase (moss.nibb.ac.jp). Every EST sequence thus obtained was examined for sequence similarity and the presence of the CHS signature sequences (the active site C, H, N triad, and the GFGPG loop), and grouped according to both UTR and coding sequences. Similarly, BLASTn and BLASTp were performed with the nucleotide and amino acid sequences of PpCHS as queries against the *P. patens* genome (*Physcomitrella patens v.1.0*) available from the DOE Joint Genome Institute (JGI, http://shake.jgi-psf.org/Phypa1/). The GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) database in the JGI site were also term-searched for annotated CHS. Because annotation of *P. patens* genes was not completed at the time of this study, each sequence obtained from these searches was examined for ORF, UTRs, and intron, and compared with *P. patens* CHS EST sequences.

5.8. Phylogenetic analysis

The amino acid sequences of 35 representative plant CHS superfamily enzymes including the six representative *P. patens* CHS superfamily genes, and six microorganism CHS-like enzymes were aligned with ClustalX (Thompson et al., 1997). Phylogenetic analysis with the Bayesian inference method was performed using the

MrBayes program (v. 3.1.2) (Ronquist and Huelsenbeck, 2003). For amino acid data, MrBayes employs various evolutionary models including the Poisson model and Blosum62 model, where both the stationary state frequencies and the substitution rates are fixed. In this study, we chose the "mixed" amino acid model that allows model jumping between various fixed-rate amino acid models. Markov chain Monte Carlo analysis (MCMC) was performed for 1 million generations with four independent chains. The Markov chain was sampled every 100 generations. At the end of run, average standard deviation of split frequencies reached close to 0.028, and the potential scale reduction factors (PSRF) approached 1.000 for all the run parameters. These diagnostics provided supports that the state of convergence was obtained. All trees sampled before 500,000 generations were discarded (burn-in = 5000), and a consensus tree was constructed based on the remaining trees using TreeView (Page, 1996).

Acknowledgements

The authors thanks B. Akister and Dr. N. Ashton, University of Regina, for providing the plant material, and E. Barker and G. Gordon for help with genome analysis. This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the University of Regina. C.K.S. was recipient of an NSERC Undergraduate Student Research Award.

Appendix A. Supplementary data

Amino acid sequences of the putative CHS superfamily genes found in the *P. patens* genome are provided. Only those genes with known ESTs are compared. The nucleotide sequences of the closely related 3′-UTRs of the putative CHS genes are also provided. Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.phytochem.2006.09.030.

References

Ashton, N.W., Schulze, A., Hall, P., Bandurski, R.S., 1985. Estimation of indole-3-acetic acid in gametophytes of the moss, *Physcomitrella patens*. Planta 164, 142–144.

Austin, M.B., Noel, J.P., 2003. The chalcone synthase superfamily of type III polyketide synthases. Nat. Prod. Rep. 20, 79–110.

Bangera, M.G., Thomashow, L.S., 1999. Identification and characterization of a gene cluster for synthesis of the polyketide antibiotic 2,4-diacetylphloroglucinol from *Pseudomonas fluorescens* Q2-87. J. Bacteriol. 181, 3155–3163.

Basile, A., Sorbo, S., Lopez-Saez, J.A., Cobianchi, R.C., 2003. Effects of seven pure flavonoids from mosses on germination and growth of *Tortula muralis* HEDW (Bryophyta) and *Raphanus sativus* L (Magnoliophyta). Phytochemistry 62, 1145–1151.

Brinkmeier, E., Geiger, H., Zinsmeister, H.D., 1999. Biflavonoids and 4,2′-epoxy-3-phenylcoumarins from the moss *Mnium hornum*. Phytochemistry 52, 297–302.

Ferrer, J.L., Jez, J.M., Bowman, M.E., Dixon, R.A., Noel, J.P., 1999. Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. Nat. Struct. Biol. 6, 775–784.

Fischer, S., Böttcher, U., Reuber, S., Anhalt, S., Weissenböck, G., 1995. Chalcone synthase in the liverwort *Marchantia polymorpha*. Phytochemistry 95, 1007–1012.

Funa, N., Ohnishi, Y., Fujii, I., Shibuya, M., Ebizuka, Y., Horinouchi, S., 1999. A new pathway for polyketide synthesis in microorganisms. Nature 400, 897–899.

Geiger, H., Markham, K.R., 1992. Campylopusaurone, an auronoflavanone biflavonoid from the mosses *Campylopus clavatus* and *Campylopus holomitrium*. Phytochemistry 31, 4325–4328.

Gross, F., Luniak, N., Perlova, O., Gaitatzis, N., Jenke-Kodama, H., Gerth, K., Gottschalk, D., Dittmann, E., Muller, R., 2006. Bacterial type III polyketide synthases: phylogenetic analysis and potential for the production of novel secondary metabolites by heterologous expression in pseudomonads. Arch. Microbiol. 185, 28–38.

Harashima, S., Takano, H., Ono, K., Takio, S., 2004. Chalcone synthase-like gene in the liverwort, *Marchantia paleacea* var *diptera*. Plant Cell Rep. 23, 167–173.

Ito, M., Ichinose, Y., Kato, H., Shiraishi, T., Yamada, T., 1997. Molecular evolution and functional relevance of the chalcone synthase genes of pea. Mol. Gen. Genet. 255, 28–37.

Jez, J.M., Bowman, M.E., Noel, J.P., 2002. Expanding the biosynthetic repertoire of plant type III polyketide synthases by altering starter molecule specificity. Proc. Natl. Acad. Sci. USA 99, 5319–5324.

Jez, J.M., Noel, J.P., 2000. Mechanism of chalcone synthase. pK_a of the catalytic cysteine and the role of the conserved histidine in a plant polyketide synthase. J. Biol. Chem. 275, 39640–39646.

Klahre, U., Hemmings-Mieszczak, M., Filipowicz, W., 1995. Extreme heterogeneity of polyadenylation sites in mRNAs encoding chloroplast RNA-binding proteins in *Nicotiana plumbaginifolia*. Plant Mol. Biol. 28, 569–574.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. 5, 150–163.

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.

Loke, J.C., Stahlberg, E.A., Strenski, D.G., Haas, B.J., Wood, P.C., Li, Q.Q., 2005. Compilation of mRNA polyadenylation signals in Arabidopsis revealed a new signal element and potential secondary structures. Plant Physiol. 138, 1457–1468.

Lütcke, H.A., Chow, K.C., Mickel, F.S., Moss, K.A., Kern, H.F., Scheele, G.A., 1987. Selection of AUG codons differs in plants and animals. EMBO J. 6, 43–48.

Magnotta, S.M., Gogarten, J.P., 2002. Multi site polyadenylation and transcriptional response to stress of a vacuolar type H^+-ATPase subunit A gene in *Arabidopsis thaliana*. BMC Plant Biol. 2, 3.

Matsumura, H., Watanabe, S., Harada, K., Senda, M., Akada, S., Kawasaki, S., Dubouzet, E.G., Minaka, N., Takahashi, R., 2005. Molecular linkage mapping and phylogeny of the chalcone synthase multigene family in soybean. Theor. Appl. Genet. 110, 1203–1209.

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.

Page, R.D.M., 1996. TreeView: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12, 357–358.

Pfeifer, V., Nicholson, G.J., Ries, J., Recktenwald, J., Schefer, A.B., Shawky, R.M., Schroder, J., Wohlleben, W., Pelzer, S., 2001. A polyketide synthase in glycopeptide biosynthesis: the biosynthesis of the non-proteinogenic amino acid (S)-3,5-dihydroxyphenylglycine. J. Biol. Chem. 276, 38370–38377.

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

Rothnie, H.M., 1996. Plant mRNA 3′-end formation. Plant Mol. Biol. 32, 43–61.

Sawant, S.V., Kiran, K., Singh, P.K., Tuli, R., 2001. Sequence architecture downstream of the initiator codon enhances gene expression and protein stability in plants. Plant Physiol. 126, 1630–1636.

Saxena, P., Yadav, G., Mohanty, D., Gokhale, R.S., 2003. A new family of type III polyketide synthases in *Mycobacterium tuberculosis*. J. Biol. Chem. 278, 44780–44790.

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

Schröder, J., 1997. A family of plant-specific polyketide synthases: facts and predictions. Trends Plant Sci. 2, 373–378.

Seshime, Y., Juvvadi, P.R., Fujii, I., Kitamoto, K., 2005. Discovery of a novel superfamily of type III polyketide synthases in *Aspergillus oryzae*. Biochem. Biophys. Res. Commun. 331, 253–260.

Shen, B., Hutchinson, C.R., 1993. Enzymatic synthesis of a bacterial polyketide from acetyl and malonyl coenzyme A. Science 262, 1535–1540.

Stafford, H.A., 1991. Flavonoid evolution: An enzymic approach. Plant Physiol. 96, 680–685.

Stöckigt, J., Zenk, M.H., 1975. Chemical syntheses and properties of hydroxycinnamoyl-coenzyme A derivatives. Z. Naturforsch. 30, 352–358.

Suh, D.-Y., Fukuma, K., Kagami, J., Yamazaki, Y., Shibuya, M., Ebizuka, Y., Sankawa, U., 2000. Identification of amino acid residues important in the cyclization reactions of chalcone and stilbene synthases. Biochem. J. 350, 229–235.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Winkel-Shirley, B., 2001. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol. 126, 485–493.

Yamaguchi, T., Kurosaki, F., Suh, D.-Y., Sankawa, U., Nishioka, M., Akiyama, T., Shibuya, M., Ebizuka, Y., 1999. Cross-reaction of chalcone synthase and stilbene synthase overexpressed in *Escherichia coli*. FEBS Lett. 460, 457–461.

Yamazaki, Y., Suh, D.-Y., Sitthithaworn, W., Ishiguro, K., Kobayashi, Y., Shibuya, M., Ebizuka, Y., Sankawa, U., 2001. Diverse chalcone synthase superfamily enzymes from the most primitive vascular plant, *Psilotum nudum*. Planta 214, 75–84.