PpASCL, a moss ortholog of anther-specific chalcone synthase-like enzymes, is a hydroxyalkylpyrone synthase involved in an evolutionarily conserved sporopollenin biosynthesis pathway

Che C. Colpitts[1], Sung Soo Kim[2], Sarah E. Posehn[1], Christina Jepson[1], Sun Young Kim[1], Gertrud Wiedemann[3], Ralf Reski[3,4], Andrew G. H. Wee[1], Carl J. Douglas[2] and Dae-Yeon Suh[1]

[1]Department of Chemistry and Biochemistry, University of Regina, Regina, SK S4S 0A2, Canada; [2]Department of Botany, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; [3]Plant Biotechnology, Faculty of Biology, University of Freiburg, 79104 Freiburg, Germany; [4]Freiburg Institute for Advanced Studies, University of Freiburg, 79104 Freiburg, Germany

Author for correspondence:
Dae-Yeon Suh
Tel: +1 306 585 4239
Email: suhdaey@uregina.ca

Received: *6 July 2011*
Accepted: *25 July 2011*

New Phytologist (2011)
doi: 10.1111/j.1469-8137.2011.03858.x

Key words: alkylpyrone synthase, gene expression, land plant evolution, *Physcomitrella patens*, polyketide synthase, protein modeling, spore wall, sporopollenin biosynthesis.

Summary

• Sporopollenin is the main constituent of the exine layer of spore and pollen walls. Recently, several Arabidopsis genes, including polyketide synthase A (*PKSA*), which encodes an anther-specific chalcone synthase-like enzyme (ASCL), have been shown to be involved in sporopollenin biosynthesis. The genome of the moss *Physcomitrella patens* contains putative orthologs of the Arabidopsis sporopollenin biosynthesis genes.

• We analyzed available *P. patens* expressed sequence tag (EST) data for putative moss orthologs of the Arabidopsis genes of sporopollenin biosynthesis and studied the enzymatic properties and reaction mechanism of recombinant PpASCL, the *P. patens* ortholog of Arabidopsis PKSA. We also generated structure models of PpASCL and Arabidopsis PKSA to study their substrate specificity.

• *Physcomitrella patens* orthologs of Arabidopsis genes for sporopollenin biosynthesis were found to be expressed in the sporophyte generation. Similarly to Arabidopsis PKSA, PpASCL condenses hydroxy fatty acyl-CoA esters with malonyl-CoA and produces hydroxyalkyl α-pyrones that probably serve as building blocks of sporopollenin. The ASCL-specific set of Gly-Gly-Ala residues predicted by the models to be located at the floor of the putative active site is proposed to serve as the opening of an acyl-binding tunnel in ASCL.

• These results suggest that ASCL functions together with other sporophyte-specific enzymes to provide polyhydroxylated precursors of sporopollenin in a pathway common to land plants.

Introduction

In flowering plants, an important event during pollen maturation in the anther is the deposition of the pollen wall, which is necessary for pollen protection and dispersal, and pollen–stigma recognition. The exine is the tough sporophyte-derived outer layer of the pollen wall and its components are produced in the tapetum, a cell layer that surrounds the inner surface of the anther locules in which the pollen grains develop (Scott *et al.*, 2004). The main constituent of exine is sporopollenin, a polymer consisting mainly of medium- to long-chain fatty acids, with a minor component of oxygenated aromatic compounds that are thought to be derived from phenylpropanoid acids. These constituents are coupled via extensive ester and ether linkages, resulting in a polymer that is extremely resistant to degradation (Domínguez *et al.*, 1999; Scott *et al.*, 2004). Although few biochemical or genetic studies on the moss spore wall have been reported, ultrastructural studies showed that moss spore walls are composed of basic layers of the outermost perine, a separating layer, the exine, and the intine. It was also suggested that both the intine and exine of

moss spores are comparable and probably homologous to these layers in pollen grains (McClymont & Larson, 1964; Olesen & Mogensen, 1978; Brown & Lemmon, 1984), and that the major component of the exine layer of the moss spores is sporopollenin, as in flowering plants (Wellman, 2004).

Recently, molecular genetic and biochemical studies have begun to reveal the biochemical pathways leading to sporopollenin synthesis. In Arabidopsis, several genes have been implicated in sporopollenin biosynthesis. The *MALE STERILITY2* (*MS2*) gene encodes a putative fatty acid reductase and may reduce very long chain fatty acids to fatty alcohols (Doan *et al.*, 2009). The cytochrome P450 genes *CYP703A2* and *CYP704B1* encode fatty acid hydroxylases that catalyze in-chain and ω-hydroxylation, respectively, of mid- to long-chain fatty acids (Morant *et al.*, 2007; Dobritsa *et al.*, 2009). Fatty alcohols and hydroxy fatty acids produced by these gene products may serve as building blocks and provide oxygen atoms for ester and ether linkages in sporopollenin. *ACOS5*, encoding a fatty acyl-CoA synthetase (de Azevedo Souza *et al.*, 2009), and the *DRL* genes, encoding dihydroflavonol 4-reductase-like enzymes, were also shown to be integral components of the sporopollenin biosynthesis pathway (Tang *et al.*, 2009; Grienenberger *et al.*, 2010). Expression of these genes is tightly regulated spatially and temporally so that their expression is initiated at the tetrad stage and restricted to the tapetal cells and microspores at the time of exine deposition. In agreement with their roles in sporopollenin synthesis, mutations in these genes result in severe defects in exine formation.

Other key anther-expressed genes implicated in sporopollenin biosynthesis encode type III polyketide synthases (PKSs), also known as anther-specific chalcone synthase-like enzymes (ASCLs). Chalcone synthase (CHS) and other type III PKSs produce a variety of secondary metabolites in plants and microorganisms by catalyzing the condensation reactions of a starter-CoA and malonyl-CoA substrates and the cyclization reaction of the linear polyketide intermediate (Fig. 1) (Abe & Morita, 2010). ASCLs and other plant type III PKSs typically show 40–50% amino acid identity, yet they share common characteristics in gene structure and conserved signature sequences. *ASCL* genes in *Brassica napus*, *Oryza sativa* and *Silene latifolia* were found to be specifically expressed in tapetal cells at the early uninucleate microspore stages of anther development (Shen & Hsu, 1992; Hihara *et al.*, 1996; Barbacar *et al.*, 1997). Atanassov *et al.* (1998) reported the anther-specific expression of an *ASCL* gene in tobacco (*Nicotiana sylvestris*) and noted that it and other *ASCL* genes form a distinct cluster of their own in a phylogenetic reconstruction of type III PKSs. On the basis of low sequence similarity to other plant type III PKSs and the temporal and spatial expression patterns of the *ASCL* genes, these authors speculated that ASCL activity might differ from those of other type III PKSs with respect to substrate/product specificity, and that ASCLs might participate in the biosynthesis of

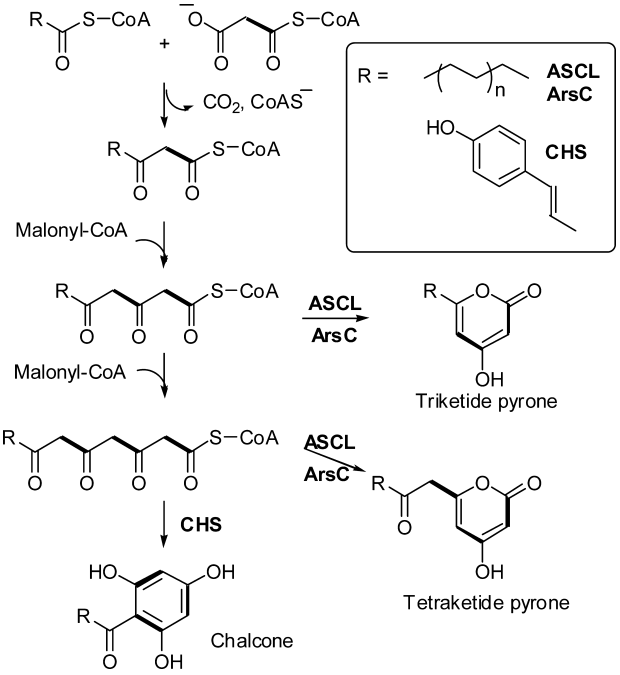


Fig. 1 Reactions catalyzed by type III polyketide synthases (PKSs). Anther-specific chalcone synthase-like enzyme (ASCL) condenses a long-chain acyl-CoA ester with two or three molecules of malonyl-CoA to produce triketide or tetraketide α-pyrone. ArsC from *Azotobacter vinelandii* also produces α-pyrone products from long-chain acyl-CoA esters. Conversely, chalcone synthase (CHS) condenses a phenylpropanoid-CoA (e.g. *p*-coumaroyl-CoA) with three molecules of malonyl-CoA and cyclizes the tetraketide intermediate to produce a chalcone.

exine. Recently, two independent studies demonstrated that mutations in the Arabidopsis *ASCL* genes *LESS ADHESIVE POLLEN (LAP6)/PKSA* (At1g02050) and *LAP5/PKSB* (At4g34850) lead to defective exine formation (Dobritsa *et al.*, 2010; Kim *et al.*, 2010). In particular, PKSA and PKSB were shown to preferentially condense hydroxy fatty acyl-CoA esters, which are produced by anther-specific fatty acid hydroxylases and ACOS, with malonyl-CoA to produce hydroxyalkyl α-pyrone compounds that probably serve as building blocks of sporopollenin (Kim *et al.*, 2010).

Comparative genomic and phylogenetic studies have demonstrated that the genome of the moss *Physcomitrella patens* contains putative orthologs of some of the sporopollenin biosynthesis genes, namely *ACOS5*, *CYP703A2*, *CYP704B1*, and *PKSA* and *PKSB* (Fig. 2) (Morant *et al.*, 2007; de Azevedo Souza *et al.*, 2009; Dobritsa *et al.*, 2009; Koduri *et al.*, 2010). This suggested that homologous genes may be involved in spore/pollen wall exine formation in mosses and spermatophytes and that biochemical pathways leading to sporopollenin biosynthesis may be conserved in land plant lineages. However, the extent of similarity in the sporopollenin biosynthesis pathways in mosses and spermatophytes and the functional orthology of the moss and spermatophyte genes remained to be investigated.

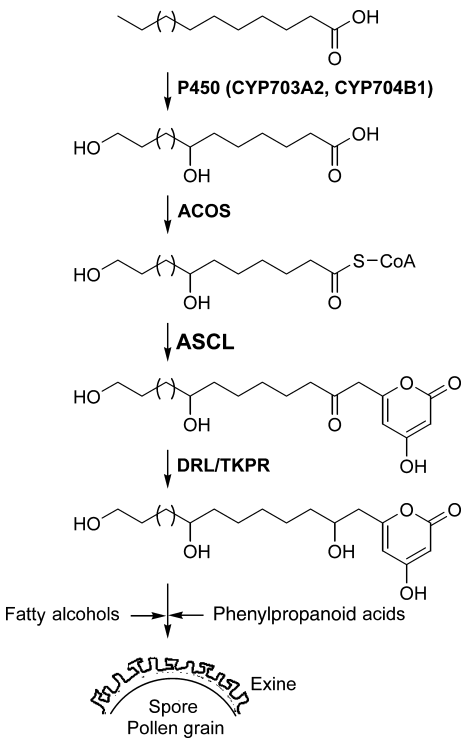


Fig. 2 Proposed role of anther-specific chalcone synthase-like enzyme (ASCL) in polyketide sporopollenin precursor biosynthesis in spore and pollen wall formation. In the proposed pathway of sporopollenin biosynthesis in developing spores or in developing anthers, ASCL condenses hydroxyl fatty acyl-CoA esters with malonyl-CoA molecules to produce tetraketide hydroxyalkyl α-pyrones. Hydroxyl fatty acyl-CoA esters are generated by sequential actions of P450 fatty acid hydroxylases and fatty acyl-CoA synthetases (ACOSs), and the pyrones are reduced by dihydroflavonol 4-reductase-like enzyme (DRL)/tetraketide α-pyrone reductase (TKPR) to polyhydroxyalkyl α-pyrones that serve as building blocks of sporopollenin.

In this study, we first identified putative *P. patens* orthologs of all of the known sporopollenin biosynthesis genes and analyzed expression patterns and expressed sequence tag (EST) abundance of the putative moss orthologs. We then compared the enzymatic activity, substrate specificity, putative substrate binding site, and reaction mechanism of recombinant PpASCL, the moss ortholog of spermatophyte ASCL, to those of Arabidopsis PKSA. The results obtained in this study provide evidence for the existence of an ancient sporopollenin biosynthetic pathway conserved in land plants, which includes the ASCL-produced α-pyrone polyketide intermediates.

Materials and Methods

Plasmids and chemicals

The EST clone (Pp020014252) containing the coding sequence (CDS) of *PpASCL* (formerly *PpCHS10*) (Koduri *et al.*, 2010) was provided by the Plant Biotechnology Department of the University of Freiburg. This clone was retrieved from a cDNA library consisting of genes expressed in sporophytes (http://www.cosmoss.org) (Lang *et al.*, 2005). Expression plasmids containing *ArsB* and *ArsC* from *Azotobacter vinelandii* (Funa *et al.*, 2006) were provided by Dr S. Horinouchi (University of Tokyo). *p*-Coumaroyl-CoA and cinnamoyl-CoA were enzymatically synthesized as described by Beuerle & Pichersky (2002). 16-Hydroxyhexadecanoyl-CoA (16-OH-C16-CoA) and 12-hydroxyoctadecanoyl-CoA (12-OH-C18-CoA), generated as described previously (Kim *et al.*, 2010), were donated by Dr E. Kombrink (Max Planck Institute for Plant Breeding Research). Triacetic acid lactone (TAL), malonyl-CoA and other acyl-CoA esters were purchased from Sigma. [2-^{14}C]Malonyl-CoA (53.9 mCi mmol^{-1}) was from PerkinElmer (Boston, MA, USA).

Methyl 3,5-dioxooctadecanoate, 3,5-dioxooctadecanoic acid and its dipotassium salt, and 4-hydroxy-6-tridecyl-2-pyrone were synthesized (Department of Chemistry and Biochemistry, University of Regina) and details of synthetic procedures and spectroscopic data of the compounds are provided in the Supporting Information Methods S1.

Identification, EST abundance and expression analysis of moss orthologs

The moss P450 enzymes belonging to the same P450 families as the Arabidopsis CYP proteins (CYP704B1, At1g69500; CYP703A2, At1g01280) were identified by examining the Cytochrome P450 database (http://drnelson.utmem.edu/cytochromeP450.html) (Nelson, 2009) and by BLASTp searches against the JGI *P. patens* genome database (http://genome.jgi-psf.org//Phypa1_1/) with the Arabidopsis enzymes as the query sequences. Similarly, two moss *DRL* orthologs were identified by BLAST searches against the *P. patens* genome using Arabidopsis *DRL1* (At4g35420) as the query sequence. The expression pattern of each putative moss ortholog was then investigated by BLASTn searches against the NCBI EST database and PHYSCObase (http://moss.nibb.ac.jp/) in order to identify the genes that are expressed in developing spores. Similarly, expression patterns of putative moss orthologs of other Arabidopsis sporopollenin biosynthesis genes were analyzed by BLASTn searches against the EST databases. Transcripts per million (TPM) data collated at the NCBI UniGene database (http://www.ncbi.nlm.nih.gov/unigene) were used to estimate the expression levels of moss orthologs.

The expression patterns of the putative moss orthologs were determined with whole genome microarrays (CombiMatrix, Mukilteo, WA, USA) based on all gene models v1.2 (http://www.cosmoss.org) (Rensing *et al.*, 2008). As starting material for RNA extraction, protonema from liquid cultures, juvenile gametophores grown on solid medium (Reski, 1998) and freshly isolated protoplasts

(Rother *et al.*, 1994) were used. The microarray experiments were performed in biological triplicates. Data analysis using the EXPRESSIONIST software (Genedata, Basel, Switzerland) was performed as previously described, using a median condensing of the probe sets and a linear array-to-array normalization with median normalization to a reference value of 10 000 (Richardt *et al.*, 2010).

Cloning, heterologous expression in *Escherichia coli* and enzyme purification

The CDS of *PpASCL* was PCR-amplified from the Pp020014252 EST clone using primers 5′-AACGA CCATGGCAAGTCGAAGGGTCGAGGCG-3′ and 5′-TGAATTCTTAGCACAGATTCCGCAGAAGAGCT-CC-3′ containing restriction sites *Nco*I and *Eco*RI, respectively (underlined). Arabidopsis total RNA was extracted from flower buds with the RNeasy Plant Mini kit (Qiagen). First-strand cDNA was generated from RNA (2.5 μg) using SuperScript II reverse transcriptase (Invitrogen), and PCR amplification of Arabidopsis *PKSA* was carried out with the primers 5′-CCATGGCTA TGTCGAACTCGAGGATGAATG-3′ and 5′-GAATT CTTAGGAAGAGGTGAGGCTGCG-3′ (*Nco*I and *Eco*RI sites underlined). The PCR products were digested with restriction enzymes and subcloned into the pET-32a(+) expression vector (Novagen), and the resulting plasmids were transformed into *E. coli* BL21(DE3) cells. The recombinant PpASCL and Arabidopsis PKSA were expressed as thioredoxin-$(His)_6$-fusion proteins and purified by Ni^{2+}-chelation chromatography as described previously (Yamazaki *et al.*, 2001).

In vitro enzyme assay and steady-state kinetic analysis

The enzyme assay was carried out as described previously (Yamazaki *et al.*, 2001). RP18 thin-layer chromatography (TLC) (Merck #1.15389) was used to separate reaction products from shorter chain and phenylpropanoid starter-CoAs (methanol/H_2O/acetic acid, 60/40/1 (v/v/v)), while silica TLC (Merck #1.11798) was used to separate products from longer chain (C_6 and up) starter-CoAs (toluene/acetone/acetic acid, 85/15/1 (v/v/v)). The radioactive products were quantified using a Storm860 phosphorimager and IMAGEQUANT (v. 5.2) software (Molecular Dynamics, Sunnyvale, CA, USA). [2-^{14}C]Malonyl-CoA was used as a standard for specific radioactivity.

The chemical identity of enzyme reaction products was verified by TLC or carrier dilution assay. The R_F (retention factor) values of the reaction products produced with acyl-CoA esters as the starter substrate were compared with those of alkylpyrones that were chemically synthesized (Methods S1) or enzymatically produced by ArsC (Funa *et al.*, 2006). The R_F values of the reaction products derived from phenylpropanoid-CoA or other CoA esters were compared with those of α-pyrones that were enzymatically prepared using *Arachis hypogaea* stilbene synthase (STS) (Suh *et al.*, 2000) or *Hydrangea macrophylla* coumaroyltriacetic acid synthase (CTAS) (Akiyama *et al.*, 1999). To further verify that PpASCL and Arabidopsis PKSA generated TAL, a triketide α-pyrone from acetyl-CoA, a carrier dilution assay was conducted as previously described (Yamaguchi *et al.*, 1999).

Kinetic experiments were conducted as described previously (Jiang *et al.*, 2006) except that the concentration of malonyl-CoA was 50 μM and the reaction time was 10 min. K_m and V_{max} were calculated by fitting the data to the Michaelis–Menten equation using GRAPHPAD PRISM v. 5.03 (GraphPad, San Diego, CA, USA).

The mechanism of pyrone formation by PpASCL and Arabidopsis PKSA

3,5-Dioxooctadecanoic acid or its dipotassium salt was incubated with Arabidopsis PKSA (20 μg) or PpASCL (100 μg) in 100 μL of 100 mM KPi buffer (pH 7.2) at 37°C for 1.5 h. The PpASCL reaction mixture contained 0.25% CHAPS (3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate) to improve the solubility of the enzyme. The reaction was terminated by acidification (15 μl of 2 N HCl). The acidification also converts the salt to the free acid form, which enables its extraction with ethyl acetate (200 μl). The ethyl acetate extracts were analyzed by silica TLC (CH_2Cl_2/methanol, 20/1).

Structure modeling of PpASCL

The structure of PpASCL was modeled initially with I-TASSER, which utilizes an *ab initio* multiple-threading approach (Zhang, 2008). The quality of the model was further improved by a 1-ns molecular dynamics simulation at 305 K in NAMD with AMBER ff99SBildn force fields. The protein was explicitly solvated during simulation in TIP3P water with Langevin dynamics and Particle Mesh Ewald (Phillips *et al.*, 2005; Lindorff-Larsen *et al.*, 2010).

Results

Moss orthologs to the Arabidopsis sporopollenin biosynthesis genes are expressed in sporophytes containing developing spores

In order to further examine the functional conservation of the sporopollenin biosynthesis genes in *P. patens* and Arabidopsis, putative *P. patens* orthologs were identified based on sequence homology and expression profile. The results are summarized in Table 1. Firstly, based on EST abundance, *PpASCL* was found to be expressed exclusively

Table 1 The *Physcomitrella patens* orthologs of the Arabidopsis genes postulated to be involved in sporopollenin biosynthesis

Enzyme	Accession number (Unigene ID)	*Physcomitrella patens* EST library (EST counts, TPM[1])	Arabidopsis ortholog	*In vitro* activity of Arabidopsis ortholog
PpASCL	XP_001781520 (Ppa.18599)	ppgs (169, 6221)	PKSA (At1g02050)[2] PKSB (At4g34850)[2] At4g00040	Type III polyketide synthase[2] Type III polyketide synthase[2]
PpACOS6[3]	XP_001767771 (Ppa.18680)	ppgs (15, 552)	At1g62940[3]	Fatty acyl-CoA synthetase[3]
PpDRL1	XP_001772000 (Ppa.18650)	ppgs (227, 8356)	At4g35420[4,5]	Tetraketide α-pyrone reductase[5]
PpDRL2	XP_001769440 (Ppa.11501)	ppgs (29, 1068)	At1g68540[5]	Tetraketide α-pyrone reductase[5]
PpCYP704B6	XP_001764503 (Ppa.18705)	ppgs (4, 147)	At1g69500[6]	CYP704B1, fatty acid ω-hydroxylase[6]
PpCYP704B7[7]	XP_001764611 (Ppa.6618)	ppgs (10, 368)	At1g69500[6]	CYP704B1, fatty acid ω-hydroxylase[6]
PpCYP703B2[8]	XP_001776045 (Ppa.10295)	ppgs (1, 37)	At1g01280[8]	CYP703A2, fatty acid in-chain hydroxylase[8]

[1]TPM, transcripts per million. Expressed sequence tag (EST) counts and TPM may be used to estimate an approximate expression level for each gene (http://www.ncbi.nlm.nih.gov/unigene/). PpDRL1 (*P. patens* dihydroflavonol 4-reductase-like) is the second most highly represented gene in the moss ppgs (green sporophytes) cDNA library, only outnumbered by *TUA4*, coding for the tubulin alpha-2/alpha-4 chain. *PpASCL* (anther-specific chalcone synthase-like) is the sixth most highly represented gene in the ppgs library. ACOS, acyl-CoA synthetase; CYP, cytochrome P450; PKS, polyketide synthase.
[2]Mizuuchi *et al.* (2008); Dobritsa *et al.* (2010); Kim *et al.* (2010); [3]de Azevedo Souza *et al.* (2009); [4]Tang *et al.* (2009); [5]Grienenberger *et al.* (2010); [6]Dobritsa *et al.* (2009); [7]Li *et al.* (2010); [8]Morant *et al.* (2007).

in the ppgs cDNA library (http://www.ncbi.nlm.nih.gov/UniGene/library.cgi?ORG=Ppa&LID=23755), which is derived from green sporophytes containing archesporial cells and, more importantly, developing spores (M. Hasebe, pers. comm.). ESTs corresponding to *PpACOS6*, the moss ortholog of Arabidopsis *ACOS5*, were also found predominantly (15 out of 17) in this library.

BLAST searches using Arabidopsis DRL1 as the query sequence yielded two *P. patens DRL* genes encoding putative ketoreductases. PpDRL1 has a higher similarity (54% amino acid sequence identity) to Arabidopsis DRL1 (At4g35420), while PpDRL2 is more similar (52% identity) to Arabidopsis DRL2 (At1g68540). ESTs for *PpDRL1* were exclusively found in the ppgs library, whereas ESTs for *PpDRL2* were predominantly (29 out of 33) found in this library. This is consistent with important roles for PpDRL1 and PpDRL2, like PpASCL and PpACOS6, in the green sporophyte stage of moss development, such as in the biosynthesis of the sporopollenin coat of the spore. Furthermore, the expression of *PpASCL* and *PpDRL1* appears to be under strict developmental regulation.

Among over 70 *P. patens P450* genes (Nelson, 2009), three *CYP703B* and two *CYP704B* homologs of the Arabidopsis anther-specific *CYP703A2* and *CYP704B1* genes were found. Earlier phylogenetic analyses suggested that *PpCYP703B2* and *PpCYP704B7* are orthologous to *CYP703A2* and *CYP704B1*, respectively (Dobritsa *et al.*, 2009; Li *et al.*, 2010). In agreement with these studies, ESTs corresponding to these two *P. patens* genes were found in the ppgs library (Table 1). In addition, ESTs corresponding to *PpCYP704B6*, which is orthologous to Arabidopsis *CYP704B1*, were also exclusively found in this library. This suggests that fatty acid oxygenation may also occur during spore development in the moss, as in angiosperms. Of the two *P. patens* genes (XP_001771307 and XP_001758118) homologous to *MS2*, one gene (XP_ 001771307; Phypa_11916) has one corresponding EST clone (ppgs36c08) in the ppgs library and none in other libraries.

The correlation of EST abundance to gene expression levels was in this case confirmed by the results of the microarray analysis, in which the probes for all transcripts are specific. In agreement with the EST abundance data which suggest exclusive expression of *PpASCL, PpDRL1* and *CYP704B6* in sporophytes, these genes are not expressed above the detection limit in protonema, gametophores and freshly isolated protoplasts. Expression of *PpCYP703B2* could be detected at a low level in the juvenile gametophores. Similar low expression levels of *PpACOS6* and *PpDRL2* were found in all three tested tissue types. The weak expression of the two putative *P. patens MS2* orthologs is also supported by the microarray analysis, as neither transcript could be seen above the detection limit (Fig. 3).

In vitro analysis of PpASCL and Arabidopsis PKSA reactions

Recombinant PpASCL and Arabidopsis PKSA were expressed as thioredoxin-fusion proteins of *c.* 60 kDa in

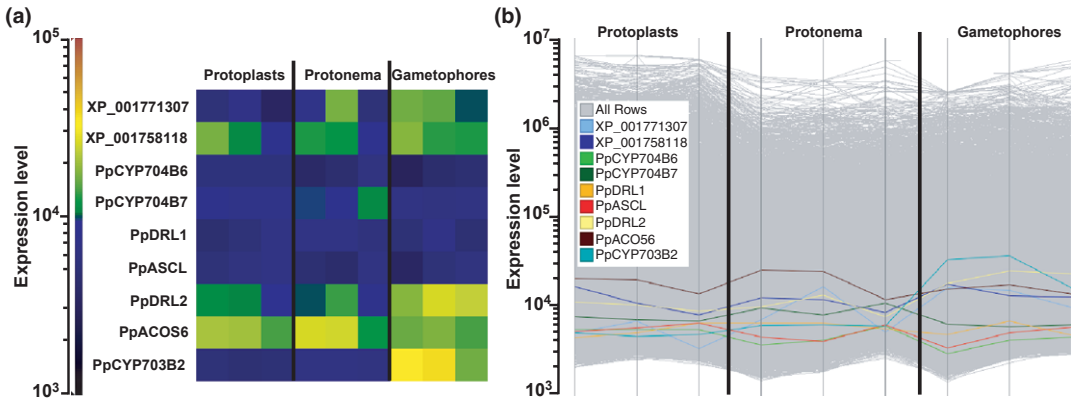
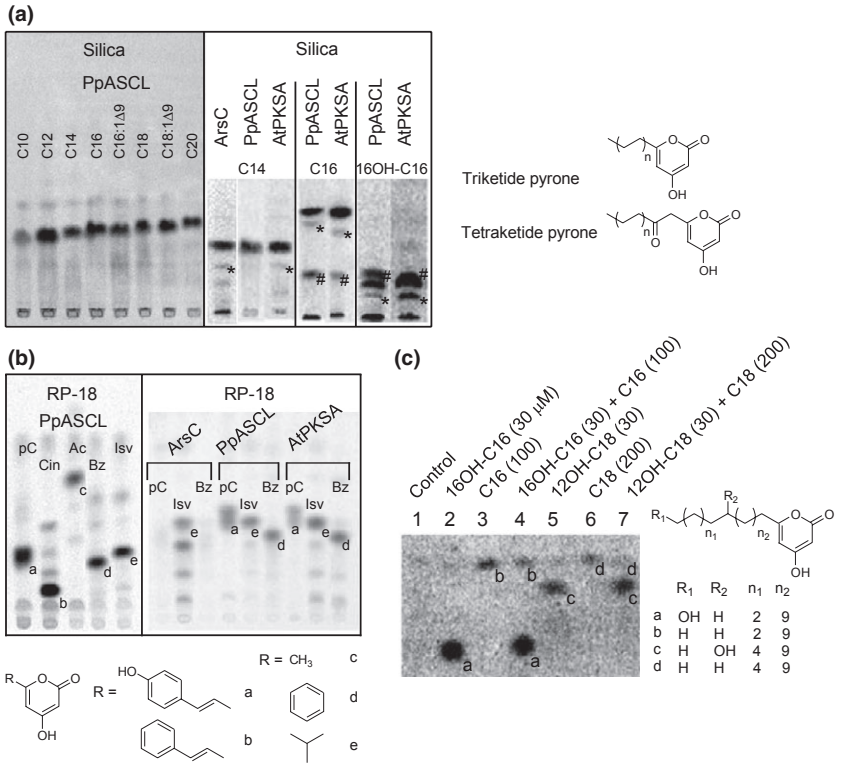


Fig. 3 Microarray analysis of the expression of sporopollenin biosynthesis genes in *Physcomitrella patens*. (a) Tile display of the expression pattern of the moss orthologs to the Arabidopsis sporopollenin biosynthesis genes and (b) profile display showing all other transcripts as gray lines in the background. Gene expression was assayed in three different tissue types in three biological replicates: protoplasts, protonema and gametophores. In (a), the color of the tiles indicates the expression level of each gene in a given tissue type. As the color approaches closer toward red, the gene is expressed at a higher level. The same data are graphically presented in (b). *PpASCL* (*P. patens* anther-specific chalcone synthase-like), *PpCYP704B6* (cytochrome P450 704B6), *PpCYP704B7* and *PpDRL1* (dihydroflavonol 4-reductase like) were not expressed in all tissues tested, while *PpDRL2*, *PpACOS6* (acyl-CoA synthetase) and *PpCYP703B2* were expressed in at least one of the tested tissues. Values below 10^4 are in the background and thus have to be considered as not expressed.

E. coli. The thioredoxin polypeptide chain was previously shown not to affect catalytic activity of type III PKSs, while improving the solubility of overproduced enzymes (Suh *et al.*, 2000). The production of tri- and tetraketide pyrones by Arabidopsis PKSA using various starter CoA esters was recently reported (Mizuuchi *et al.*, 2008; Dobritsa *et al.*, 2010; Kim *et al.*, 2010). Under the reaction conditions used in this study, PpASCL as well as Arabidopsis PKSA accepted a wider range of starter substrates than previously reported (Mizuuchi *et al.*, 2008; Dobritsa *et al.*, 2010). When a saturated (C6–C20) or unsaturated (C16:1 or C18:1) fatty acyl-CoA was used as the starter substrate, PpASCL produced a single major product. With increasing chain length of the starter substrate, the R_F value of the product also progressively increased, indicating that PpASCL produced the same type of compounds from fatty acyl CoAs of different chain lengths (Fig. 4a). The reaction products of PpASCL co-migrated on TLC with the reaction products of Arabidopsis PKSA and ArsC (Fig. 4a, right panel). ArsC was also shown to produce triketide α-pyrones as major products using various fatty acyl-CoA esters (Fig. 1) (Funa *et al.*, 2006). Furthermore, the reaction product generated using C14 myristoyl-CoA as a starter substrate co-migrated with synthetic 4-hydroxy-6-tridecyl-2-pyrone (R_F = 0.30, silica-TLC, toluene/acetone/acetic acid, 85/15/1; R_F = 0.27, CH_2Cl_2/methanol, 20/1). These results indicate that, under the reaction conditions used, PpASCL also condenses fatty acyl-CoA with two molecules of malonyl-CoA to produce triketo α-pyrone as a major product, similar to the Arabidopsis PKSA and PKSB enzymes.

When other CoA esters such as *p*-coumaroyl-CoA, cinnamoyl-CoA, acetyl-CoA, benzoyl-CoA or isovaleryl-CoA were used as starter substrates, PpASCL also produced a single major product (Fig. 4b). The chemical identity of the major products was determined by comparing their chromatographic behaviors with those of triketide α-pyrones produced by plant type III PKSs. Thus, when incubated with *p*-coumaroyl-CoA, PpASCL generated a single radioactive band of R_F 0.27 on RP-TLC, which co-migrated with the triketide α-pyrone *bis*noryangonin produced as a derailment product by STS or CTAS (Akiyama *et al.*, 1999; Suh *et al.*, 2000). Carrier dilution analysis provided additional support that the products of the PpASCL reactions were triketide α-pyrones. Three successive recrystallizations of radioactive PpASCL-derived TAL in the presence of non-labeled TAL in ethyl acetate and one more recrystallization in ethanol gave crystals of constant specific radioactivity: 5900, 5600, 5450 and 6000 dpm mg^{-1}, respectively. Arabidopsis PKSA also produced triketide α-pyrones as major products from the same phenylpropanoid-CoA and short chain acyl-CoA esters (Fig. 4b). Conversely, ArsC only accepted isovaleryl-CoA to give multiple products including triketide α-pyrone and did not utilize either *p*-coumaroyl-CoA or benzoyl-CoA (Fig. 4b, right panel).

In addition to the triketide α-pyrone, PpASCL, like Arabidopsis PKSA, produced a minor product when incubated with acyl-CoA starter substrates (Fig. 4a, right panel). The amount of the minor product produced by both enzymes varied depending on the starter substrate used; however, typically < 1% of the amount of the triketide α-pyrone was produced under the reaction conditions used. The minor products were presumed to be tetraketide α-pyrones (Fig. 1), as they showed lower R_F values than those of the corresponding triketide α-pyrones, similar to the tetraketide α-pyrones produced by Arabidopsis PKSA and ArsC (Fig. 4a, right panel) (Funa *et al.*, 2006; Mizuuchi



(d)

Substrate	K_m (μM)	k_{cat} (10^{-4} sec^{-1})	k_{cat}/K_m (M^{-1} sec^{-1})
16-OH-C16-CoA	21 ± 4.3	1.9 ± 0.19	9.0
12-OH-C18-CoA	18 ± 7.5	0.71 ± 0.084	3.9
C16-CoA	160 ± 11	0.53 ± 0.14	0.33

Fig. 4 Radio-thin-layer chromatography (TLC) of products generated by anther-specific chalcone synthase-like enzyme (PpASCL), the *Physcomitrella patens* ortholog of Arabidopsis polyketide synthase A (PKSA). In all parts of the figure, TLC spots without notation marks are those of triketide pyrone products. Minor tetraketide pyrone products are denoted by an asterisk, while the hash symbol denotes unused [2-^{14}C]malonyl-CoA molecules that were carried over to the organic phase during extraction. (a) Radio silica-TLC of reaction products produced by PpASCL from acyl-CoA starter substrates of varying chain lengths and [2-^{14}C]malonyl-CoA. The substrates used were decanoyl-CoA (C10), lauroyl-CoA (C12), myristoyl-CoA (C14), palmitoyl-CoA (C16), palmitoleoyl-CoA (C16:1Δ9), stearoyl-CoA (C18), oleoyl-CoA (C18:1Δ9) and arachidoyl-CoA (C20). Some of reaction products generated by ArsC and Arabidopsis PKSA are also shown for comparison. (b) Radio reversed-phase (RP)-TLC of reaction products produced by PpASCL, Arabidopsis PKSA and ArsC from various starter CoA esters and [2-^{14}C]malonyl-CoA. The starter-CoAs used were *p*-coumaroyl-CoA (pC), cinnamoyl-CoA (Cin), acetyl-CoA (Ac), benzoyl-CoA (Bz) and isovaleryl-CoA (Isv). (c) Radio silica-TLC of the substrate competition assay. To a PpASCL reaction mixture containing 50 μM [2-^{14}C]malonyl-CoA was added 16-OH-C16-CoA (30 μM), or C16-CoA (100 μM), or both 16-OH-C16-CoA (30 μM) and C16-CoA (100 μM). Similarly, the PpASCL reaction was carried out in the presence of 12-OH-C18-CoA at 30 μM, or C18-CoA at 200 μM, or both 12-OH-C18-CoA and C18-CoA at 30 μM and 200 μM, respectively. The control sample received [2-^{14}C]malonyl-CoA only. Each reaction produced a triketide α-pyrone as the major product (a–d). To measure competition, the levels of α-pyrone production in lane 4 (or 7) were compared to those in lanes 2 and 3 (or 5 and 6). (d) The steady-state kinetic parameters of recombinant PpASCL. Values given are mean ± SE; n = 3.

et al., 2008; Kim *et al.*, 2010). Based on these results, it was concluded that PpASCL is a moss enzyme functionally orthologous to Arabidopsis PKSA.

Substrate preference and steady-state kinetics of PpASCL

Arabidopsis PKSA was recently shown to prefer hydroxy fatty acyl CoA esters (16-OH-C16-CoA and 12-OH-C18-CoA) as the starter substrate relative to unsubstituted fatty acyl-CoAs (Kim *et al.*, 2010). Similar substrate competition assays showed that PpASCL also prefers hydroxy fatty acyl-CoA esters. When the enzyme reaction was carried out in the presence of 16-OH-C16-CoA at 30 μM and C16-CoA at 100 μM, the production of 15′-OH-C15-α-pyrone (the triketide product from 16-OH-C16-CoA) was reduced to 80% as compared with the control reaction with 16-OH-C16-CoA as the sole substrate (2.5 vs 2.0 pmol of 15′-OH-

C15-α-pyrone) (Fig. 4c, lanes 2 and 4). Conversely, the production of C15-α-pyrone (the triketide product from C16-CoA) was reduced to 51% of the control level (0.55 vs 0.28 pmol of C15-α-pyrone) (lanes 3 and 4). Substrate preference was more evident in the pair of 12-OH-C18-CoA and C18-CoA. While C18-CoA at 200 μM had little effect on the production of 17-OH-C17-α-pyrone (lanes 5 and 7), 12-OH-C18-CoA at 30 μM reduced the production of C17-α-pyrone to 46% of the control level (0.24 vs 0.11 pmol of C17-α-pyrone) (lanes 6 and 7).

Enzyme kinetics provided further evidence for the substrate preference (Figs 4d, S1). PpASCL preferred 16-OH-C16-CoA to C16-CoA with lower K_m and higher k_{cat} values. Thus, the k_{cat}/K_m values for 16-OH-C16-CoA and 12-OH-C18-CoA (9.0 and 3.9 $M^{-1}s^{-1}$, respectively) are 27- and 12-fold higher than the k_{cat}/K_m for C16-CoA (0.33 $M^{-1}s^{-1}$). These results strongly suggest that hydroxy fatty acyl-CoA esters, which are expected to be generated by a moss ortholog of Arabidopsis ACOS5, are probably the preferred *in planta* substrates of PpASCL, as suggested for Arabidopsis PKSA (Kim *et al.*, 2010).

The mechanism of pyrone formation by ASCLs

To determine if di- (or tri-)keto acid is an intermediate in the *in vitro* formation of the tri- (or tetra-)ketide α-pyrone product of ASCLs (pathway ***B*** in Fig. 5), a synthetic diketo acid was incubated with the enzymes. After incubation of 3,5-dioxooctadecanoic acid or its dipotassium salt at different concentrations of up to 1 mM with Arabidopsis PKSA or PpASCL, the diketo acid was recovered unchanged (R_F = 0.17 (streaking), CH_2Cl_2 : methanol, 20 : 1). Even after incubation at pH 2, the diketo acid remained unchanged and no trace of the triketo α-pyrone (4-hydroxy-6-tridecyl-2-pyrone) (R_F = 0.27) was detected on TLC. A representative thin-layer chromatogram is shown in Fig. S2. Therefore, the diketo acid, added either as a free acid or as a salt, was not converted to the corresponding triketo α-pyrone either enzymatically or spontaneously. These results indicate that the di- (or tri-)keto acid is not an intermediate in the pyrone formation by ASCLs and, by elimination, support pathway ***A*** in which ASCLs produce α-pyrones by direct *O*-acylation of oligoketide-CoA thioester intermediates.

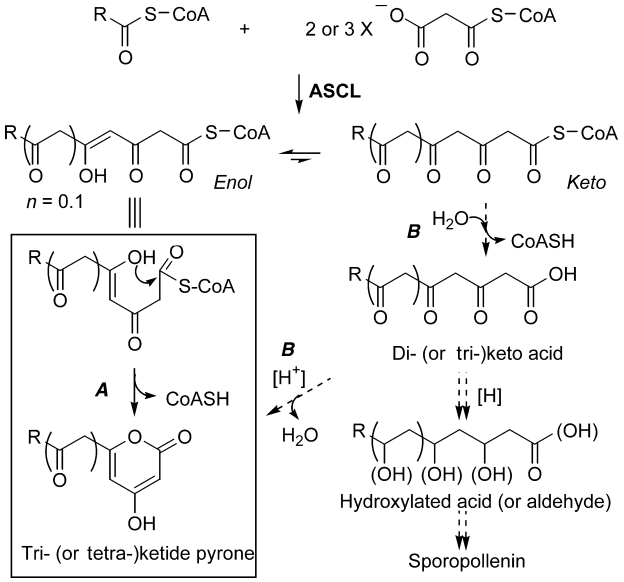


Fig. 5 The mechanism of α-pyrone formation by anther-specific chalcone synthase-like enzyme (ASCL). Pathway ***A*** involves the nucleophilic attack of the enolic oxygen on the thioester carbon. Pathway ***B*** involves the hydrolysis of the oligoketide CoA thioester to oligoketo acid, which presumably undergoes acid-catalyzed ring formation and dehydration. The double-dashed arrows denote unlikely pathways in which ASCLs produce oligoketo acids *in planta*, which might subsequently be reduced to hydroxylated acids or aldehydes as sporopollenin precursors.

Structure modeling and active site architecture of PpASCL and Arabidopsis PKSA

Despite low (< 50%) overall sequence identity, the overall three-dimensional structure of I-TASSER-generated models of PpASCL and Arabidopsis PKSA closely resembled those of MsCHS and other plant type III PKSs (Figs 6a, S3a). Amino acid residues that have been shown to be important for substrate binding and catalysis are conserved in ASCLs. In the refined PpASCL model, these conserved residues are found at similar positions when compared with the known type III PKS structures; namely, the catalytic triad of Cys^{192}, His^{331} and Asn^{364} at the putative active site, and Phe^{243}, Phe^{293} and the $A^{402}FGPG$ loop at the putative cyclization pocket (Fig. 6b) (Ferrer *et al.*, 1999; Jez *et al.*, 2000b; Suh *et al.*, 2000).

Comparison of putative active site residues of ASCLs and non-ASCL enzymes suggested clues as to the different enzyme activity of ASCLs. First, a total of 26 active site residues of MsCHS were identified that are within 5 Å of the bound naringenin in the co-crystal structure of MsCHS and naringenin (PDB id, 1cgk). The corresponding putative active site residues of PpASCL and Arabidopsis PKSA were then identified from sequence alignment and the modeled structures (Figs 6B, S3b and S4). When the sequences of 14 ASCLs (Koduri *et al.*, 2010) and non-ASCLs were compared, these active site residues were found to be conserved or conservatively substituted in both ASCLs and non-ASCLs except that a Gly (Gly^{225} in PpASCL and Gly^{205} in Arabidopsis PKSA) in ASCLs is substituted with a bulkier residue such as Thr in MsCHS (Thr^{197}) or Leu in *Gerbera hybrida* 2-pyrone synthase (2PS) (Fig. 7).

Among the 13 amino acid residues that are uniquely conserved in ASCLs (Fig. S4), Ala^{240} and Gly^{239} are found side by side with the ASCL-specific Gly^{225} on the floor of the putative active site of PpASCL (Fig. 6b). In MsCHS, the corresponding residue to Ala^{240} is Gln^{212} (Fig. 7). This Gln

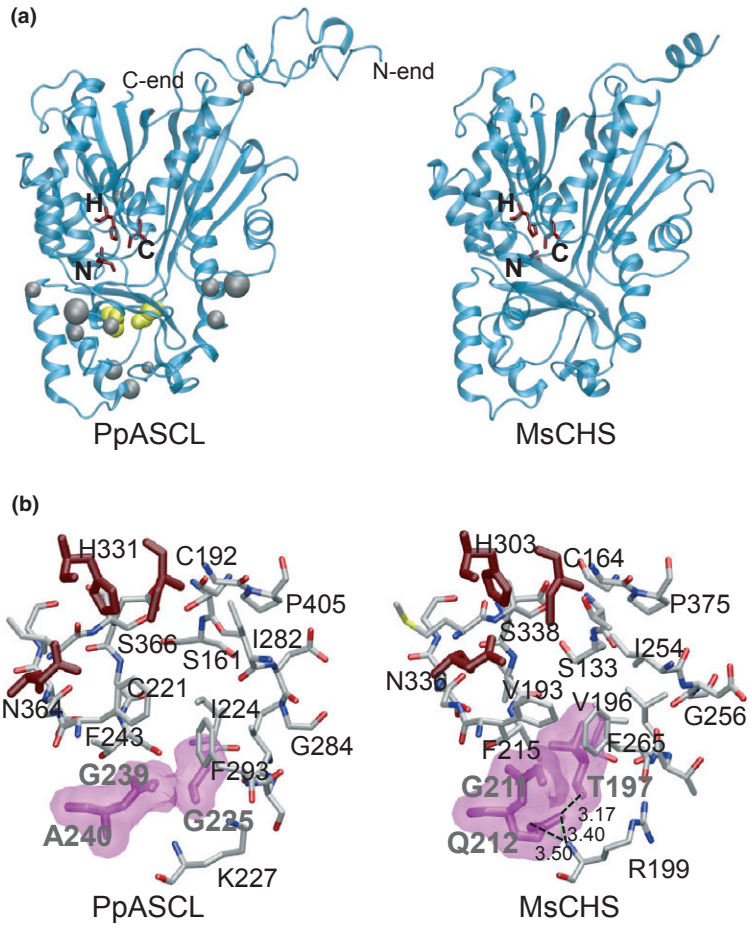


Fig. 6 Comparison of I-TASSER-generated model of PpASCL (*P. patens* anther-specific chalcone synthase-like enzyme) and the crystal structure of *Medicago sativa* chalcone synthase (MsCHS). (a) The I-TASSER-generated PpASCL model (left) and the crystal structure of MsCHS (1bi5, right) are shown. The modeled PpASCL structure was found to be close to the X-ray crystal structure quality according to the parameters of MolProbity (Davis *et al.*, 2007), with the structure found in the 99th percentile of overall quality compared with all known crystal structures. The scores for Ramachandran outliers (0.96%), poor rotamers (2.08%) and clash score (0.16) were roughly the same as the average of those for X-ray crystal structures of plant type III polyketide synthases (PKSs). The Cys-His-Asn catalytic triads are shown in brick red in both structures. Shown in yellow space-filling models are Gly^{225} and Ala^{240} of PpASCL that are on the active site floor and are proposed to form the opening of the acyl-binding tunnel. The gray dots indicate ASCL-specific residues that are not modeled to be at the active site (Supporting Information Fig. S4). The three-dimensional perspective of residues is indicated by different dot sizes. (b) The modeled active site of PpASCL (left) is compared with the active site of MsCHS (right). The Gly^{225}-Gly^{239}-Ala^{240} set at the putative tunnel entrance of PpASCL is shown as a magenta space-filling model with the corresponding Thr^{197}-Gly^{211}-Gln^{212} set in MsCHS. H-bonds among the side chain of Q^{212} of MsCHS and neighboring residues are indicated by lines. Numbers are interatomic distances in Å. The catalytic triads are in brick red as in (a). Oxygen, nitrogen and sulfur atoms are rendered red, blue and yellow, respectively.

residue is highly conserved in non-ASCLs. In the crystal structure of MsCHS, the amide O_ε of Gln^{212} is in H-bond distance from the backbone N (3.50 Å) of Arg^{199} and its amide N_ε is in H-bond distances from the backbone O of Thr^{197} (3.17 Å) and from the backbone N of Arg^{199} (3.40 Å) (Fig. 5b). The bulkiness and ability to form multiple H-bonds of the Gln^{212} side chain seem to contribute to sealing the floor of the active site and limit the size of acceptable substrates for MsCHS. Conversely, in the modeled structure of PpASCL and Arabidopsis PKSA, Ala^{240} (Ala^{220} of PKSA) is neighbored (5–6 Å) by other nonpolar residues of Tyr^{69} (Phe^{49} of PKSA) and Tyr^{226} (Phe^{206} of PKSA). Smaller sizes of Ala^{240} and nearby Gly^{225} and their inability to form multiple H-bonds make these two residues primary candidates to be situated at the opening of the acyl-binding tunnel in PpASCL (Fig. 6b). A similar argument is put forward for Arabidopsis PKSA, in which the corresponding residues at the putative tunnel opening are Gly^{205} Gly^{219} and Ala^{220} (Fig. S3b).

Discussion

Evolution of the sporopollenin biosynthesis pathway

This study provides the first biochemical evidence that one of the moss orthologs (PpASCL) of Arabidopsis sporopollenin

PpASCL	^{160}SSeirL	^{191}GC	^{220}ECTlIGy	^{239}GA	^{243}FGDG
AtPKSA	^{140}SSeirL	^{171}GC	^{200}ETTiLGf	^{219}GA	^{223}FGDG
MsCHS	^{132}TSgvdM	^{163}GC	^{192}EVTaVTf	^{211}GQ	^{215}FGDG
Gh2PS	^{137}TAgvdM	^{168}GC	^{197}EITaILf	^{216}AQ	^{220}FGDG
RpALS	^{133}TTsndM	^{164}GC	^{193}EIVaFAf	^{212}GQ	^{216}FGDG
AaPCS	^{142}SCgvdM	^{173}GC	^{202}ELTiMMl	^{221}GI	^{225}FGDG
AaOKS	^{142}SCgvdM	^{173}GC	^{202}ELTiIGl	^{221}GN	^{225}FGDG
PKS18	^{143}STgfiA	^{174}GC	^{203}ELCsVNa	^{220}IH	^{224}FGDG
ORAS	^{120}CTdsaN	^{151}GC	^{184}EVStTMv	^{206}GI	^{210}FSDC

Fig. 7 Sequence alignment of active site residues. Active site residues of *Medicago sativa* chalcone synthase (MsCHS) and their neighboring residues are aligned with corresponding residues of PpASCL (*P. patens* anther-specific chalcone synthase-like enzyme) and other type III polyketide synthases (PKSs). Gly and Ala residues (Gly^{225} and Ala^{240}) that are proposed to form the putative acyl-binding tunnel entrance of PpASCL are highlighted in gray The active site nucleophile Cys residues are underlined and non-active site residues are in lowercase letters. For brevity, active site residues that are not close to the proposed tunnel opening are not shown. The complete sequence alignment is shown in Supporting Information Fig. S4. AaPCS, *Aloe arborescens* pentaketide chromone synthase (AAX35541); AaOKS, *Aloe arborescens* octaketide synthase (AAT48709); AtPKSA, Arabidopsis polyketide synthase A (NP_171707); Gh2PS, *Gerbera hybrida* 2-pyrone synthase (P48391); ORAS, *Neurospora crassa* 2′-oxoalkylresorcylic acid synthase (XP_960427); PKS18, *Mycobacterium tuberculosis* PKS18 (YP_177803); RpALS, *Rheum palmatum* aloesone synthase (AAS87170).

biosynthesis genes indeed carries out the same function *in vitro* as the Arabidopsis counterpart (PKSA). Our finding that a set of Arabidopsis genes encoding sporopollenin biosynthetic enzymes, including ASCLs, are conserved in *P. patens* and are specifically expressed in developing moss sporophytes further supports the existence of an ancient sporopollenin biosynthetic pathway conserved in land plants, which includes the hydroxyalkylpyrone intermediates.

Sporopollenin is probably not a unique macromolecule but rather a mixture of related biopolymers. Although the lack of detailed chemical studies means that caution should be employed in interpreting the data, the presence of compounds described as sporopollenin has been reported in several algal species, including *Coleochaete* (Guilford *et al.*, 1988; Delwiche *et al.*, 1989; Ueno, 2009), in fruiting bodies of the cellular slime molds (Maeda, 1984), and in several myxobacteria species (Strohl *et al.*, 1977), as well as in land plants. In view of its ancient origin and unparalleled chemical stability, it is generally accepted that sporopollenin was important in the first land plants for protection of spores from UV irradiation, physical damage, and microbial attack during the successful colonization of land (Wellman, 2004).

The strategy of incorporating type III PKS-generated long-chain hydroxyalkylpyrones into protective structures is not restricted to plant lineages. Alkylpyrones and alkylresorcinols produced by type III PKSs were shown to be essential to cyst formation and membrane integrity in microorganisms (Funa *et al.*, 2006; Funabashi *et al.*, 2008). It has been proposed that sporopollenin encloses resting cysts or reproductive structures of various algal groups (Wellman, 2004, and references therein). However, little is known about the chemical composition, structure, and synthesis of algal sporopollenin, and whether it is homologous to the sporopollenin found in land plants. The emerging genetic information on the land plant sporopollenin biosynthesis pathway should enable comparative genomics approaches to investigate the pathway in algae and its relationship to the land plant polymer.

PpASCL, similar to Arabidopsis PKSA, is a hydroxyalkylpyrone synthase with substrate preference for hydroxylfatty acyl-CoA esters

The observed substrate preferences of PpASCL (Fig. 4c) and Arabidopsis PKSA (Kim *et al.*, 2010) for hydroxylfatty acyl-CoA esters support the hypothesis that ASCLs utilize hydroxy fatty acyl-CoA esters generated by the sequential actions of P450s and a fatty acyl-CoA synthetase to generate sporopollenin precursors, and that this role is conserved in land plant lineages including moss and Arabidopsis.

Plant type III PKSs are well known for their substrate promiscuity (Abe & Morita, 2010), and PpASCL and Arabidopsis PKSA are no exception. Even so, the range of accepted substrates of these enzymes is unusual as no other type III PKSs have been shown to utilize phenylpropanoid-CoAs and (unsubstituted) long-chain acyl-CoAs. However, as the *in planta* activities of the ASCLs will be governed by substrate availability, the substrate promiscuity is probably not physiologically significant. Key enzymes in the biosynthetic pathway for phenylpropanoid-CoAs are phenylalanine ammonia-lyase (PAL) and 4-coumarate:coenzyme A ligase (4CL). Although the *P. patens* genome has ten *PAL* and four *4CL* genes, we failed to find EST sequences corresponding to any *PAL* and *4CL* genes in the ppgs library derived from sporophytic material. Further, we confirmed the very basal expression of these genes in gametophytic material by microarray analyses (data not shown). Taken together, these results indicate that phenylpropanoid-CoA esters are not available to PpASCL *in planta*. Similarly, Arabidopsis *ACOS5* is strongly co-expressed with Arabidopsis *PKSA* in tapetum cells of developing anthers (de Azevedo Souza *et al.*, 2009), but there is no evidence for *4CL* expression in the tapetum,

suggesting that ACOS5-derived fatty acyl-CoAs are the major substrate available to Arabidopsis PKSA *in planta* (Kim *et al.*, 2010). Thus, it seems that there has not been strong evolutionary pressure to refine the enzyme substrate-binding site of ASCLs so as to exclude possible competitor substrates. This argument is consistent with the lack of chalcones/flavonoids in sporopollenin (Boavida *et al.*, 2005), as these compounds would require *p*-coumaroyl-CoA as a starter substrate.

ASCL may produce tetraketide α-pyrones *in planta*

As α-pyrones are produced by most type III PKSs as *in vitro* derailment products (Yamaguchi *et al.*, 1999), the *in planta* relevance of the tri- and tetraketide α-pyrones produced by PpASCL and other ASCLs must be considered. Two possibilities for the *in planta* products of these enzymes are (1) PpASCL produces tetraketide α-pyrones as the major (or sole) product *in planta*. In this case, triketide α-pyrones observed *in vitro* would be derailment products. (2) PpASCL catalyzes CHS- or STS-type cyclization under physiological conditions, and both triketide and tetraketide α-pyrones are *in vitro* derailment products as a consequence of suboptimal reaction conditions used *in vitro*. The second possibility is less likely, as ArsB, another long-chain acyl-CoA accepting enzyme (Funa *et al.*, 2006), successfully catalyzed cyclization and produced alkylresorcinols under the same reaction conditions (data not shown). We failed to detect any trace of 5-pentadecylresorcinol after PpASCL or Arabidopsis PKSA was incubated with C16-CoA and malonyl-CoA. Conversely, the observation that Arabidopsis DRL1, a ketoreductase shown to be essential in exine formation (Tang *et al.*, 2009; Grienenberger *et al.*, 2010), specifically accepts the tetraketide α-pyrone produced by Arabidopsis PKSA and reduces the acyl ketone group, while lacking detectable activity against triketide α-pyrone, strongly suggests that the physiological function of Arabidopsis PKSA and other ASCLs is to generate tetraketide α-pyrones that are further modified by downstream enzymes, including DRL1, before being incorporated into sporopollenin (Grienenberger *et al.*, 2010). DRL has been renamed tetraketide α-pyrone reductase (TKPR) to reflect its catalytic action on tetraketide pyrones (Grienenberger *et al.*, 2010). Based on these results, a biochemical pathway for sporopollenin biosynthesis is proposed (Fig. 2). According to this model, ASCL converts hydroxy fatty acyl-CoA esters, which are produced by P450 and ACOS, to tetraketide α-pyrones. The pyrones are then reduced by DRL/TKPR to polyhydroxyalkyl α-pyrones that serve as building blocks of sporopollenin.

The possibility that ASCLs might produce oligoketide-CoA thioesters or oligoketo acids *in planta*, and pass them to downstream enzymes for further modification (e.g. reduction to hydroxylate acids or aldehydes; Fig. 5) was also considered. Such an example exists in lignin biosynthesis: cinnamoyl-CoA reductase reduces cinnamoyl-CoA to cinnamyl aldehyde in monolignol biosynthesis (Lacombe *et al.*, 1997). Furthermore, although it is generally accepted that type III PKSs produce α-pyrones by *O*-acylation of oligoketide-CoA thioester (pathway ***A***; Fig. 5) (Jez *et al.*, 2000a; Funa *et al.*, 2006), an alternate pathway involving hydrolysis followed by nonenzymatic pyrone formation during acid work-up (pathway ***B***; Fig. 5) was proposed for the derailment pyrone production in type III PKS reactions (Schüz *et al.*, 1983; Suh *et al.*, 2000). Our finding that diketo acid was not converted either enzymatically or spontaneously to α-pyrone indicates that pathway ***B*** does not represent the mechanism of pyrone formation by ASCLs and also suggests that keto acids are not sporopollenin precursors produced by ASCLs *in planta*. It is noted that our results do not exclude the possibility, albeit theoretical, that ASCL-produced oligoketide-CoA thioesters are metabolically channelled to a yet-to-be-identified downstream enzyme for further modification.

The conserved Gly and Ala residues may form the opening of the acyl-binding tunnel in ASCL

It has been repeatedly demonstrated that the starter substrate specificity of type III PKSs and the length of polyketide products are modulated by subtle changes in active site cavity, sometimes by single amino acid substitutions (Abe & Morita, 2010). Possibly as a result of lower sequence homology between ASCLs and other type III PKSs, homology-based modeling, which has been successfully employed for other type III PKSs (Abe & Morita, 2010), did not result in high-quality models for PpASCL and Arabidopsis PKSA. Instead, we obtained structure models of desirable quality for both enzymes using *ab initio* multiple-threading and molecular dynamics simulation methods. From the comparison of the modeled PpASCL active site and the MsCHS active site, it is clear that the most prominent difference is at the region of Gly^{225}, Gly^{239} and Ala^{240} (Fig. 6b). In MsCHS, these residues are substituted with Thr^{197}, Gly^{211} and Gln^{212}. Thr^{197} and corresponding residues in type III PKSs were shown to play important roles in substrate specificity and chain length determination in 2PS (Jez *et al.*, 2000a), octaketide synthase (OKS) (Abe *et al.*, 2005a), pentaketide chromone synthase (PCS) (Abe *et al.*, 2005b), and aloesone synthase (Abe *et al.*, 2004) (Fig. 7). For example, the point mutation of Gly^{207} of OKS, a PKS that performs seven sequential condensations, to a bulkier residue resulted in the formation of various shorter chain length polyketide compounds including a triketide, TAL (Abe *et al.*, 2005a). Conversely, the single mutation of Met^{207} of PCS to Gly transformed the pentaketide-forming PCS into an octaketide-producing enzyme (Abe *et al.*, 2005b). These findings suggest that the Thr-to-Gly substitution in ASCLs may play a similar role in expanding the active site cavity in the enzyme.

CHS, STS, and most other non-ASCL-type III PKSs have a Gln in place of Ala^{240} (Fig. 7). The side chain of Gln^{212} in MsCHS forms multiple H-bonds with its neighboring chain, supporting the floor of the active site (see Results). Similar H-bonds between the conserved Gln and neighboring chain are also observed in 2PS (PDB id, 1ee0). It seems reasonable, then, to suggest that the conserved Gln in CHS and other non-ASCLs plays a role in restricting the size of the active site cavity, thereby controlling the size of acceptable starter substrates. In contrast, the smaller size of Ala and its inability to form strong dipole interactions may be critical in providing structural flexibility necessary for tunnel opening in ASCLs. It is noted that some ASCLs have a Val in place of the Ala (Fig. S4).

In summary, the presence of orthologs of most Arabidopsis genes of sporopollenin biosynthesis in the *P. patens* genome, and their specific expression in the sporophyte generation, strongly suggest that the biosynthesis pathway of sporopollenin is well conserved in land plants, spanning an evolutionary distance of *c.* 500 million yr (Lang *et al.*, 2008). PpASCL is the functional ortholog of ASCL in that both enzymes are hydroxyalkylpyrone synthases that prefer hydroxy fatty acyl-CoA esters as substrates. Conserved Gly-Gly-Ala residues are proposed to form the opening of the acyl-binding tunnel specific to the ASCL active site. Because of the unique position of bryophytes in land plant evolution and the fact that *P. patens* is amenable to targeted genetic manipulation (Hohe *et al.*, 2004; Khraiwesh *et al.*, 2008), the study of sporopollenin biosynthesis in the moss should provide valuable insights into the biosynthesis and evolution of sporopollenin, a biopolymer of interest as a consequence of its extreme resistance to degradation.

Acknowledgements

This research was supported by the Natural Sciences and Engineering Research Council (NSERC) Discovery grants to D.-Y.S., A.G.H.W. and C.J.D. C.C.C. thanks NSERC for a CGS-M postgraduate scholarship. We thank M. Vervliet-Scheebaum and S. A. Rensing (University of Freiburg) for *PpASCL* EST clone, and S. Horinouchi (University of Tokyo) for expression plasmids of *ArsB* and *ArsC*. We also thank T. Beuerle (TU Braunschweig) for a *4CL* expression plasmid.

References

Abe I, Morita H. 2010. Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases. *Natural Product Reports* **27**: 809–838.

Abe I, Oguro S, Utsumi Y, Sano Y, Noguchi H. 2005a. Engineered biosynthesis of plant polyketides: chain length control in an octaketide-producing plant type III polyketide synthase. *Journal of American Chemical Society* **127**: 12709–12716.

Abe I, Utsumi Y, Oguro S, Morita H, Sano Y, Noguchi H. 2005b. A plant type III polyketide synthase that produces pentaketide chromone. *Journal of American Chemical Society* **127**: 1362–1363.

Abe I, Watanabe T, Lou W, Noguchi H. 2004. Active site residues governing substrate selectivity and polyketide chain length in aloesone synthase. *FEBS Journal* **273**: 208–218.

Akiyama T, Shibuya M, Liu HM, Ebizuka Y. 1999. *p*-Coumaroyltriacetic acid synthase, a new homologue of chalcone synthase, from *Hydrangea macrophylla* var. *thunbergii*. *European Journal of Biochemistry* **263**: 834–839.

Atanassov I, Russinova E, Antonov L, Atanassov A. 1998. Expression of an anther-specific chalcone synthase-like gene is correlated with uninucleate microspore development in *Nicotiana sylvestris*. *Plant Molecular Biology* **38**: 1169–1178.

de Azevedo Souza C, Kim SS, Koch S, Kienow L, Schneider K, McKim SM, Haughn GW, Kombrink E, Douglas CJ. 2009. A novel fatty acyl-CoA synthetase is required for pollen development and sporopollenin biosynthesis in *Arabidopsis*. *Plant Cell* **21**: 507–525.

Barbacar N, Hinnisdaels S, Farbos I, Monéger F, Lardon A, Delichère C, Mouras A, Negrutiu I. 1997. Isolation of early genes expressed in reproductive organs of the dioecious white campion (*Silene latifolia*) by subtraction cloning using an asexual mutant. *Plant Journal* **12**: 805–817.

Beuerle T, Pichersky E. 2002. Enzymatic synthesis and purification of aromatic coenzyme A esters. *Analytical Biochemistry* **302**: 305–312.

Boavida LC, Becker JD, Feijó JA. 2005. The making of gametes in higher plants. *International Journal of Developmental Biology* **49**: 595–614.

Brown RC, Lemmon BE. 1984. Spore wall development in *Andreaea* (Musci: Andreaeopsida). *American Journal of Botany* **71**: 412–420.

Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB III, Snoeyink J, Richardson JS *et al.* 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research* **35**: W375–W383.

Delwiche CF, Graham LE, Thomson N. 1989. Lignin-like compounds and sporopollenin in *Coleochaete*, an algal model for land plant ancestry. *Science* **245**: 399–401.

Doan TT, Carlsson AS, Hamberg M, Bulow L, Stymne S, Olsson P. 2009. Functional expression of five *Arabidopsis* fatty acyl-CoA reductase genes in *Escherichia coli*. *Journal of Plant Physiology* **166**: 787–796.

Dobritsa AA, Lei Z, Nishikawa S, Urbanczyk-Wochniak E, Huhman DV, Preuss D, Sumner LW. 2010. *LAP5* and *LAP6* encode anther-specific proteins with similarity to chalcone synthase essential for pollen exine development in *Arabidopsis thaliana*. *Plant Physiology* **153**: 937–955.

Dobritsa AA, Shrestha J, Morant M, Pinot F, Matsuno M, Swanson R, Møller BL, Preuss D. 2009. CYP704B1 is a long-chain fatty acid ω-hydroxylase essential for sporopollenin synthesis in pollen of *Arabidopsis*. *Plant Physiology* **151**: 574–589.

Domínguez E, Mercado JA, Quesada MA, Heredia A. 1999. Pollen sporopollenin: degradation and structural elucidation. *Sexual Plant Reproduction* **12**: 171–178.

Ferrer JL, Jez JM, Bowman ME, Dixon RA, Noel JP. 1999. Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nature Structural Biology* **6**: 775–784.

Funa N, Ozawa H, Hirata A, Horinouchi S. 2006. Phenolic lipid synthesis by type III polyketide synthases is essential for cyst formation in *Azotobacter vinelandii*. *Proceedings of the National Academy of Sciences, USA* **103**: 6356–6361.

Funabashi M, Funa N, Horinouchi S. 2008. Phenolic lipids synthesized by type III polyketide synthase confer penicillin resistance on *Streptomyces griseus*. *Journal of Biological Chemistry* **283**: 13983–13991.

Grienenberger E, Kim SS, Lallemand B, Geoffroy P, Heintz D, de Azevedo Souza C, Heitz T, Douglas CJ, Legrand M. 2010. Analysis of *TETRAKETIDE α-PYRONE REDUCTASE* function in *Arabidopsis thaliana* reveals a previously unknown, but conserved, biochemical

pathway in sporopollenin monomer biosynthesis. *Plant Cell* **22**: 4067–4083.

Guilford WJ, Schneider DM, Labovitz J, Opella SJ. 1988. High resolution solid state ^{13}C NMR spectroscopy of sporopollenins from different plant taxa. *Plant Physiology* **86**: 134–136.

Hihara Y, Hara C, Uchimiya H. 1996. Isolation and characterization of two cDNA clones for mRNAs that are abundantly expressed in immature anthers of rice (*Oryza sativa* L.). *Plant Molecular Biology* **30**: 1181–1193.

Hohe A, Egener T, Lucht JM, Holtorf H, Reinhard C, Schween G, Reski R. 2004. An improved and highly standardised transformation procedure allows efficient production of single and multiple targeted gene-knockouts in a moss, *Physcomitrella patens*. *Current Genetics* **44**: 339–347.

Jez JM, Austin MB, Ferrer J, Bowman ME, Schröder J, Noel JP. 2000a. Structural control of polyketide formation in plant-specific polyketide synthases. *Chemistry & Biology* **7**: 919–930.

Jez JM, Ferrer JL, Bowman ME, Dixon RA, Noel JP. 2000b. Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase. *Biochemistry* **39**: 890–902.

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

Khraiwesh B, Ossowski S, Weigel D, Reski R, Frank W. 2008. Specific gene silencing by artificial microRNAs in *Physcomitrella patens*: an alternative to targeted gene knockouts. *Plant Physiology* **148**: 684–693.

Kim SS, Grienenberger E, Lallemand B, Colpitts CC, Kim SY, de Azevedo Souza C, Geoffroy P, Heintz D, Krahn D, Kaiser M *et al.* 2010. *LAP6/ POLYKETIDE SYNTHASE A* and *LAP5/ POLYKETIDE SYNTHASE B* encode hydroxyalkyl α-pyrone synthases required for pollen development and sporopollenin biosynthesis in *Arabidopsis thaliana*. *Plant Cell* **22**: 4045–4066.

Koduri PKH, Gordon GS, Barker EI, Colpitts CC, Ashton NW, Suh D-Y. 2010. Genome-wide analysis of the chalcone synthase superfamily genes of *Physcomitrella patens*. *Plant Molecular Biology* **72**: 247–263.

Lacombe E, Hawkins S, Van Doorsselaere J, Piquemal J, Goffner D, Poeydomenge O, Boudet AM, Grima-Pettenati J. 1997. Cinnamoyl CoA reductase, the first committed enzyme of the lignin branch biosynthetic pathway: cloning, expression and phylogenetic relationships. *Plant Journal* **11**: 429–441.

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

Lang D, Zimmer AD, Rensing SA, Reski R. 2008. Exploring plant biodiversity: the *Physcomitrella* genome and beyond. *Trends in Plant Science* **13**: 542–549.

Li H, Pinot F, Sauveplane V, Werck-Reichhart D, Diehl P, Schreiber L, Franke R, Zhang P, Chen L, Gao Y *et al.* 2010. Cytochrome P450 family member CYP704B2 catalyzes the ω-hydroxylation of fatty acids and is required for anther cutin biosynthesis and pollen exine formation in rice. *Plant Cell* **22**: 173–190.

Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**: 1950–1958.

Maeda Y. 1984. The presence and location of sporopollenin in fruiting bodies of the cellular slime moulds. *Journal of Cell Science* **66**: 297–308.

McClymont JW, Larson DA. 1964. An electron-microscopic study of spore wall structure in the Musci. *American Journal of Botany* **51**: 195–200.

Mizuuchi Y, Shimokawa Y, Wanibuchi K, Noguchi H, Abe I. 2008. Structure function analysis of novel type III polyketide synthases from *Arabidopsis thaliana*. *Biological & Pharmaceutical Bulletin* **31**: 2205–2210.

Morant M, Jorgensen K, Schaller H, Pinot F, Moller BL, Werck-Reichhart D, Bak S. 2007. CYP703 is an ancient cytochrome P450 in land plants catalyzing in-chain hydroxylation of lauric acid to provide building blocks for sporopollenin synthesis in pollen. *Plant Cell* **19**: 1473–1487.

Nelson DR. 2009. The cytochrome P450 homepage. *Human Genomics* **4**: 59–65.

Olesen P, Mogensen GS. 1978. Ultrastructure, histochemistry and notes on germination stages of spores in selected mosses. *Bryologist* **81**: 493–516.

Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. 2005. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26**: 1781–1802.

Rensing SA, Lang D, Zimmer A, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud P-F, Lindquist E, Kamisugi Y *et al.* 2008. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* **319**: 64–69.

Reski R. 1998. Development, genetics and molecular biology of mosses. *Botanica Acta* **111**: 1–15.

Richardt S, Timmerhaus G, Lang D, Qudeimat E, Correa LGG, Reski R, Rensing SA, Frank W. 2010. Microarray analysis of the moss *Physcomitrella patens* reveals conserved transcriptional regulation of salt stress and abscisic acid signalling. *Plant Molecular Biology* **72**: 27–45.

Rother S, Hadeler B, Orsini JM, Abel WO, Reski R. 1994. Fate of a mutant macrochloroplast in somatic hybrids. *Journal of Plant Physiology* **143**: 72–77.

Schüz R, Heller W, Hahlbrock K. 1983. Substrate specificity of chalcone synthase from *Petroselinum hortense*. Formation of phloroglucinol derivatives from aliphatic substrates. *Journal of Biological Chemistry* **258**: 6730–6734.

Scott RJ, Spielman M, Dickinson HG. 2004. Stamen structure and function. *Plant Cell* **16**: S46–S60.

Shen JB, Hsu FC. 1992. *Brassica* anther-specific genes: characterization and *in situ* localization of expression. *Molecular Genetics and Genomics* **234**: 379–389.

Strohl W, Larkin JM, Good BH, Chapman RL. 1977. Isolation of sporopollenin from four myxobacteria. *Canadian Journal of Microbiology* **23**: 1080–1083.

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. *Biochemical Journal* **350**: 229–235.

Tang LK, Chu H, Yip WK, Yeung EC, Lo C. 2009. An anther-specific dihydroflavonol 4-reductase-like gene (*DRL1*) is essential for male fertility in *Arabidopsis*. *New Phytologist* **181**: 576–587.

Ueno R. 2009. Visualization of sporopollenin-containing pathogenic green micro-alga *Prototheca wickerhamii* by fluorescent *in situ* hybridization (FISH). *Canadian Journal of Microbiology* **55**: 465–472.

Wellman CH. 2004. Origin, function and development of the spore wall in early land plants. In: Hemsley AR, Poole I, eds. *The evolution of plant physiology: from whole plants to ecosystems*. London, UK: Elsevier, 43–64.

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 Letters* **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.

Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**: 40.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Initial velocity plots for the measurement of kinetic parameters of *Physcomitrella patens* anther-specific chalcone synthase-like enzyme (PpASCL).

Fig. S2 Thin-layer chromatogram of the incubation mixture of a diketo acid and Arabidopsis polyketide synthase A (PKSA).

Fig. S3 I-TASSER-generated model of Arabidopsis polyketide synthase A (PKSA).

Fig. S4 Comparison of deduced amino acid sequences of anther-specific chalcone synthase-like enzymes (ASCLs).

Methods S1 Chemical syntheses.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

Supporting Information

Methods S1. Chemical Syntheses

Methyl 3,5-dioxooctadecanoate (**1**) was synthesized following a literature method (Lygo, 1995). The dianion of methyl acetoacetate (NaH, BuLi, THF, 0°C) was acylated with *N*-tetradecanoyl-2-methylaziridine, which after 1 h was quenched with 10% aq. HCl. The crude product was subjected to flash chromatographic purification using acid-washed silica gel to yield **1** as a mixture of enol and keto tautomers. The ratio of the enol to keto-form was determined to be ~9:1, based on the integration of the enol H-6 singlet at 2.29 ppm and the keto methylene H-6 singlet at 2.50 ppm.

Methyl 3,5-dioxooctadecanoate (**1**): IR (CH_2Cl_2): ν 1743, 1603 cm^{-1}. ^1H NMR (Signals of minor keto tautomer in square brackets) δ ($CDCl_3$, 300 MHz): 0.88 (t, 3H, *J* = 6.5 Hz, H-18), 1.18−1.39 (m, 20H, H-8−17), 1.55−1.65 (m, 2H, H-7), 2.29 ("t") and [2.50, "t"] (2H, *J* = 7.4 Hz, H-6), 3.35 (s) and [3.57, s], (2H, H-2), 3.74 (s, 3H, OMe), 5.59 (s, enol-H-4) and [3.60, s, keto-H-4], and 15.13 (br s, enol-OH) (2H). HRMS calcd for $C_{19}H_{34}O_4$ 326.2457, found 326.2459.

The diketo ester **1** was treated with KOH (5.5 mol eq) in absolute ethanol at −20°C for 20 h (Schmidt *et al*., 2006) during which time the dipotassium salt **2** precipitated out of solution. The salt **2** was redissolved in water and the solution was acidified to pH 2 with aq. 1 M HCl. The resultant carboxylic acid was extracted with cold $CHCl_3$ and the combined extracts were dried (anhydrous Na_2SO_4), filtered and concentrated to give 3,5-dioxooctadecanoic acid **3** as a white solid, m.p. 73.0−73.6°C. ^1H NMR δ ($CDCl_3$, 300 MHz): 0.88 (t, 3H, *J* = 6.9 Hz, H-18), 1.15−1.41 (m, 20H, H-8–17), 1.51−1.68 (m, 2H, H-7), 2.30 ("t", 2H, *J* = 7.4 Hz, H-6), 3.41 (s, 2H, H-2), 5.58 (s, 1H, H-4), 14.77 (br s, 1H, enol-OH).

Trifluoroacetic acid (2.2 mmol) was added to a suspension of **2** (1 mmol) in trifluoroacetic acid anhydride at −20°C. The resulting solution was allowed to warm to 0°C and was stirred for 2 h. The mixture was poured into ice water, extracted with CH_2Cl_2, dried over Na_2SO_4, filtered and concentrated to give the crude product. Purification via flash column chromatography (CH_2Cl_2/methanol, 20/1, R_F = 0.27) gave the target pyrone.

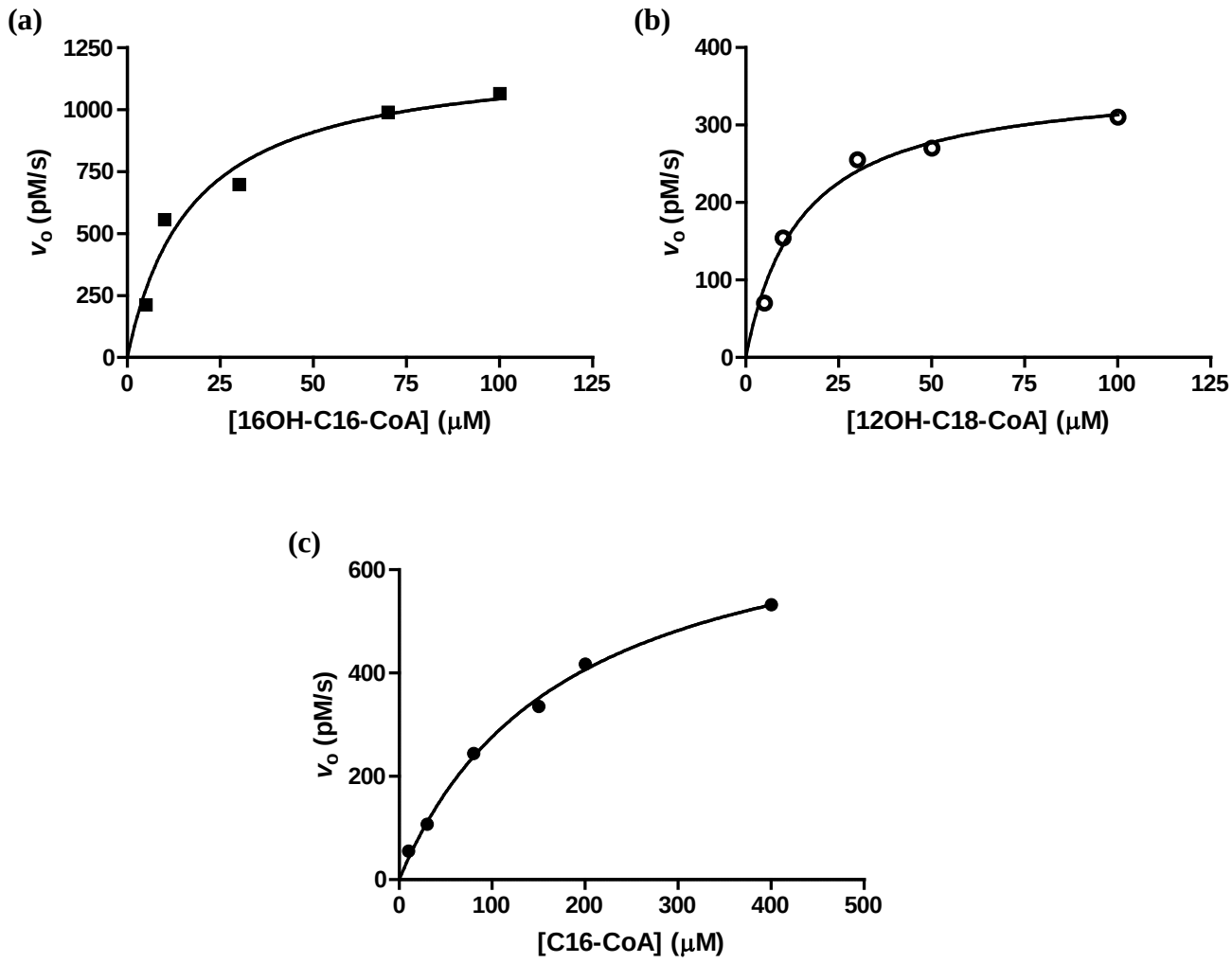
4-Hydroxy-6-tridecyl-2-pyrone: ^1H NMR δ ($CDCl_3$, 300 MHz): 0.88 (t, 3H, H-13′), 1.19−1.39 (m, 20H, H-3′−12′), 1.55−1.70 (m, 2H, H-2′), 2.47 ("t", 2H, *J* = 7.5 Hz, H-1′), 5.52 (d, 1H, ^4J = 1.8 Hz, H-3), 5.92 (d, 1H, ^4J = 1.8 Hz, H-5), 9.25 (br s, 1H, enol-OH).

References

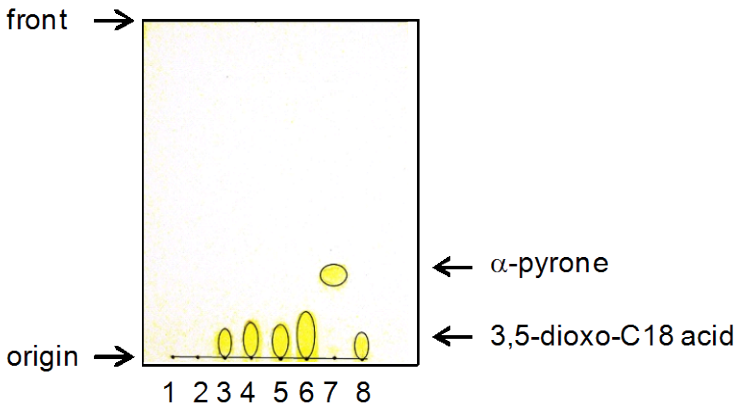
Lygo B. 1995. *N*-Acyl-2-methylaziridines: Synthesis and utility in the *C*-acylation of β-ketoester derived dianions. *Tetrahedron* **51:** 12859−12868.

Schmidt D, Conrad J, Klaiber I, Beifuss U. 2006. Synthesis of the bis-potassium salts of 5-hydroxy-3-oxopent-4-enoic acids and their use for the efficient preparation of 4-hydroxy-2*H*-pyran-2-ones and other heterocycles. *Chemical Communication*s 4732−4734.

Supporting Information Fig. S1. Initial velocity plots for the measurement of kinetic parameters of PpASCL for different substrates, 16-OH-C16-CoA (a), 12-OH-C18-CoA (b), and C16-CoA (c). The experimental data were fit to the Michaelis-Menten equation by nonlinear regression curve-fitting method.

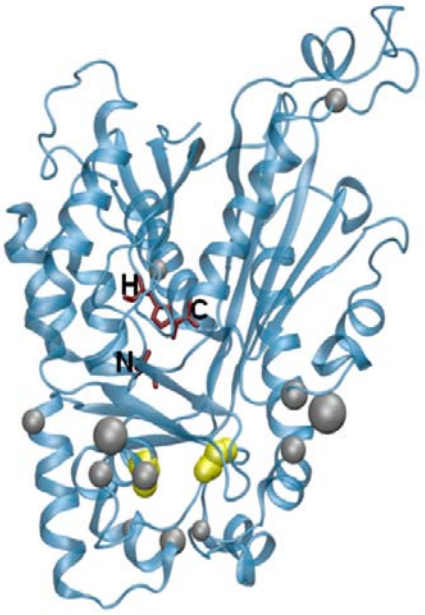


Supporting Information Fig. S2. Thin-layer chromatogram of the incubation mixture of a diketo acid and Arabidopsis polyketide synthase A (PKSA).
Dipotassium salt of 3,5-dioxooctadecanoic acid **2** was incubated with *Arabidopsis* PKSA (20 µg) in 100 µl of 100 mM KPi buffer (pH 7.2) at 37°C for 1.5 h. The reaction was terminated by addition of 15 µl of 2 N HCl in order to convert the remaining salt to the corresponding acid. The reaction mixture was extracted with ethyl acetate (200 µL) and the extracts were analyzed by silica TLC (CH_2Cl_2/methanol, 20/1). The compounds were stained with iodide vapour. Lane 1, buffer only; lane 2, *Arabidopsis* PKSA in the buffer; lane 3, PKSA plus **2** (200 µM); lane 4, **2** (200 µM) only without PKSA; lane 5, PKSA plus **2** (1 mM); lane 6, **2** only (1 mM) without PKSA; lane 7, synthetic 4-hydroxy-6-tridecyl-2-pyrone, lane 8: synthetic 3,5-dioxooctadecanoic acid.

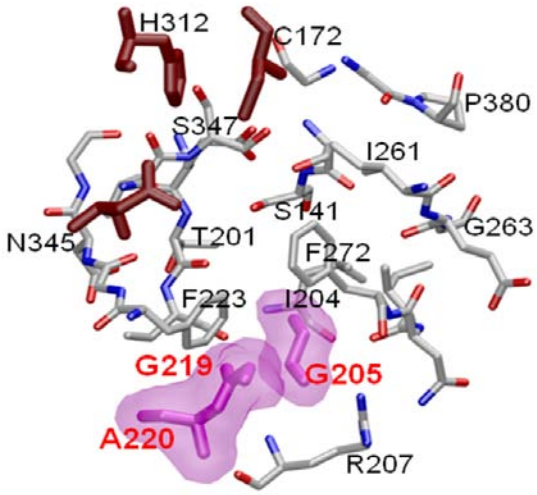


Supporting Information Fig. S3. I-TASSER-generated model of Arabidopsis polyketide synthase A (PKSA).
(a) The Cys-His-Asn catalytic triad is shown in brick red. Gly^{205} and Ala^{220} that are proposed to form the entrance of acyl-binding tunnel are shown in yellow space filling model. Other ASCL-specific residues that are not modeled to be at the active site (supplemental Fig. 3) are indicated with gray dots. The size of dot reflects 3-D perspective. The modeled structure had the Morbidity score of 1.42, which puts the structure in the 97^{th} percentile of overall quality compared with all known crystal structures. The scores for Ramachandran outliers (1.02%), poor rotamers (4.24%) and clash score (0) were comparable to those of the PpASCL model.
(b) The modeled putative active site of *Arabidopsis* PKSA. The Gly^{205}-Gly^{219}-Ala^{220} set at the putative tunnel entrance are shown in space-filling model. The catalytic triad is shown in brick red as in (a).

(a)



(b)



Supporting Information Fig. S4. Comparison of deduced amino acid sequences of anther-specific chalcone synthase-like enzymes (ASCLs).
Active site amino acid residues of MsCHS and the corresponding residues of ASCLs and two other representative non-ASCL enzymes are highlighted in cyan. The active site residues of MsCHS are defined as those within 5 Å of naringenin bound at the active site cavity (PDB id, 1cgk). The T to G substitution and the Q to A (or V) substitution in ASCLs are highlighted in yellow. Other ASCL-specific residues that are not modeled to be at the putative active site are indicated in gray.

Enzyme	Species	Taxa	Accession No.
ASCLs			
PpASCL	*Physcomitrella patens*	Bryophyta, Funariales	XP_001781520
PrCHSL	*Pinus radiata*	Spermatophyta, Coniferales	AAB80804
Os10g0484800 (YY2)	*Oryza sativa*	Liliopsida, Poales	NP_001064891
NsCHSLK	*Nicotiana sylvestris*	asterids, Solanales	CAA74847
PKSA	*Arabidopsis thaliana*	eurosids II, Brassicales	NP_171707,
At1g02050			
PKSB			NP_567971,
At4g34850			
AtCHSL2			NP_191915,
At4g00040			
AhCHSL1	*Arabidopsis halleri*	eurosids II, Brassicales	AAZ81872
AhCHSL2			AAZ23686
ZmCHSL	*Zea mays*	Liliopsida, Poales	NP_001149508
HvCHSL	*Hordeum vulgare*	Liliopsida, Poales	AAV49989
TaCHSL			CAJ15412
SlCHSL	*Silene latifolia*	Caryophyllales, Caryophyllaceae	BAE80096
VvCHSL	*Vitis vinifera*	Vitales, Vitaceae	CAO47307
Non-ASCLs			
MsCHS	*Medicago sativa*	eurosids I, Fabales	P30074
AhSTS	*Arachis hypogaea*	eurosids I, Fabales	P20178
Gh2PS	*Gerbera* hybrid cultivar	asterids, Asterales	P48391

```
TaCHSL         ----------------MVSARGVDTTTAANKQQQATCLAPNPGKATILALGHAFPQQLV
HvCHSL         ----------------MVSARDVDTNKQQQQQQATCLLAPNPGKATILALGHAFPQQLV
ZmCHSL         ----------------MVSS---SMDTTSD-KRASSMLAPNPGKATILALGHAFPQQLV
PKSB           -------------------------MGSIDAAVLGSEKKSNPGKATILALGKAFPHQLV
AhCHSL2        -------------------------MGSIDAEVLGSEKKSNPGKATILALGKAFPHQLV
VvCHSL         -------------------------MGSGTVEQVGSLKKANPGKATILALGKAFPHQLV
SlCHSL         -------------------------MGFENIKLNGMGKKPTPGKATVLSLGKGFPHTLV
Os10g0484800   -------------------MADLG-FGDARSGNGSRSQCS-RGKAMLLALGKGLPEQVL
PKSA           -------------------MSNSRMNGVEKLSSKSTRRVANAGKATLLALGKAFPSQVV
AhCHSL1        -------------------MSNSRMNGVEKLSSISTRRVANPRKATLLALGKAFPSQVV
AtCHSL2        -------------------MLVSAR--VEKQ-----KRVAYQGKATVLALGKALPSNVV
NsCHSLK        ---------------------------------------------------MGKAFPAQLV
PrCHSL         ----------------------MSASNGTNGVVAVKSRRQHRPGKTTAMAFGRAFPDQLV
PpASCL         MASRRVEAAFDGQAVELGATIPAANGNGTHQSIKVPGHRQVTPGKTTIMAIGRAVPANTT 60

MsCHS          ---------------------------MVSVSEIRKAQRAEGPATILAIGTANPANCV 31
AhSTS          ---------------------------MVSVSGIRKVQRAEGPATVLAIGTANPPNCV
Gh2PS          ----------------------MGSYSSDDVEVIREAGRAQGLATILAIGTATPPNCV 36
```

```
TaCHSL          MQDYVVEGFMRNTNCND-PELKEKLARLCKTTTVKTRYVVMSDEILKSYPELAQEGLPTMK
HvCHSL          MQDYVVEGFMRNTNCND-PELKEKLTRLCKTTTVKTRYVVMSEEILKSYPELAQEGLPTMK
ZmCHSL          MQDYVVDGFMKNTNCDD-PELKEKLTRLCKTTTVRTRYVVMSDEILKNYPELAQEGLPTMN
PKSB            MQEYLVDGYFKTTKCDD-PELKQKLTRLCKTTTVKTRYVVMSEEILKKYPELAIEGGSTVT
AhCHSL2         MQEYLVDGYFKTTKCDD-PELKQKLTRLCKTTTVKTRYVVMSEEILKKYPELAIEGGSTVT
VvCHSL          MQEFLVDGYFRNTNCDD-PDLKEKLARLCKTTTVKTRYVVMSEEILRKYPELVIEGQPTVK
SlCHSL          MQEFLVDGYFRNTNCDD-PELKQKLTRLCKTTTVKTRYVVMSDEILKKCPELAMAGQATVK
Os10g0484800    PQEKVVETYLQDTICDD-PATRAKLERLCKTTTVRTRYTVMSKELLDEHPELRTEGTPTLT
PKSA            PQENLVEGFLRDTKCDD-AFIKEKLEHLCKTTTVKTRYTVLTREILAKYPELTTEGSPTIK
AhCHSL1         PQENLVEGFLRDTKCDD-AFIKEKLEHLCKTTTVKTRYTVLSREILDKYPELTTEGSPTIK
AtCHSL2         SQENLVEEYLREIKCDN-LSIKDKLQHLCKSTTVKTRYTVMSRETLHKYPELATEGSPTIK
NsCHSLK         PQDCLVEGYIRDTNCQD-LAIKEKLERLCKTTTVKTRYTVMSKEILDKYPELATEGTPTIK
PrCHSL          MQEFLVDGYFRNTNCQD-PVLRQKLERLCKTTTVKTRYVVMSDEILAQHPELAVEGSATVR
PpASCL          FNDGLADHYIQEFNLQD-PVLQAKLRRLCETTTVKTRYLVVNKEILDEHPEFLVDGAATVS 120

MsCHS           EQSTYPDFYFKITNSEHKTELKEKFQRMCDKSMIKRRYMYLTEEILKENPNVCEYMAPSLD  92
AhSTS           DQSTYADYYFRVTNGEHMTDLKKKFQRICERTQIKNRHMYLTEEILKENPNMCAYKAPSLD
Gh2PS           AQADYADYYFRVTKSEHMVDLKEKFKRICEKTAIKKRYLALTEDYLQENPTMCEFMAPSLN  97



TaCHSL          QRLDISNKAVTQMATEASLACVKAWGGDLSAITHLVYVSSSEARFPGGDLHLARALGLSP
HvCHSL          QRLDISNKAVTQMATEASLACVEAWGGDLSAITHLVYVSSSEARFPGGDLHLARALGLSP
ZmCHSL          QRLDISNAAVTQMATEASLSCVRSWGGALSSITHLVYVSSSEARFPGGDLHLARALGLSP
PKSB            QRLDICNDAVTEMAVEASRACIKNWGRSISDITHVVYVSSSEARLPGGDLYLAKGLGLSP
AhCHSL2         QRLDICNDAVTEMAVEASRACIKNWGRSISEITHLVYVSSSEARLPGGDLYLAKGLGLSP
VvCHSL          QRLDICNKAVTQMAIDASKACIKKWGRSVSEITHLVYVSSSEARLPGGDLYLAKGLGLSP
SlCHSL          QRLDICNDAVTEMAIDASKACISDWGRPISDITHLVYVSSSEARLPGGDLYLAKGLGLSP
Os10g0484800    PRLDICNAAVLELGATAARAALGEWGRPAADITHLVYISSSELRLPGGDLFLATRLGLHP
PKSA            QRLEIANEAVVEMALEASLGCIKEWGRPVEDITHIVYVSSSEIRLPGGDLYLSAKLGLRN
AhCHSL1         QRLEIANEAVVEMALEASLGCIKEWGRPVEDITHIVYVSSSEIRLPGGDLYLSAKLGLRN
AtCHSL2         QRLEIANDAVVQMAYEASLVCIKEWGRAVEDITHLVYVSSSEFRLPGGDLYLSAQLGLSN
NsCHSLK         QRLEIANPAVVEMAKQASQACIKEWGRSAEEITHIVYVSSSEIRLPGGDLYLATELGLRN
PrCHSL          QRLEISNVAVTDMAVDACRDCLKEWGRPVSEITHLVYVSSSEIRLPGGDLYLASRLGLRS
PpASCL          QRLAITGEAVTQLGHEAATAAIKEWGRPASEITHLVYVSSSEIRLPGGDLYLAQLLGLRS 180

MsCHS           ARQDMVVVEVPRLGKEAAVKAIKEWGQPKSKITHLIVCTTSGVDMPGADYQLTKLLGLRP 152
AhSTS           AREDMMIREVPRVGKEAATKAIKEWGQPMSKITHLIFCTTSGVALPGVDYELIVLLGLDP
Gh2PS           ARQDLVVTGVPMLGKEAAVKAIDEWGLPKSKITHLIFCTTAGVDMPGADYQLVKLLGLSP 157



TaCHSL          DVRRVMLAFTGCSGGVAGLRVAKGLAESC-PGARVLLATSETTVAGFRPPSPDRPYDLVGV
HvCHSL          DVRRVMLAFTGCSGGVAGLRVAKGLAESC-PGARVLLATSETTVAGFRPPSPDRPYDLVGV
ZmCHSL          DVRRVMLAFTGCSGGVAGLRVAKGLAESC-PGARVLLATSETTIVGFRPPSPDRPYDLVGV
PKSB            DTHRVLLYFVGCSGGVAGLRVAKDIAENN-PGSRVLLATSETTIIGFKPPSVDRPYDLVGV
AhCHSL2         DTHRVLLYFVGCSGGVAGLRVAKDIAENN-PGSRVLLATSETTIIGFKPPSVDRPYDLVGV
VvCHSL          ETHRVQLYFMGCSGGVAGLRVAKDIAENN-PESRVLLATSETTIIGFKPPSADRPYDLVGV
SlCHSL          ETNRVMLYFSGCSGGVAGFRVAKDIAENN-PGSRVLLATSETTIIGFKPPSADRPYDLVGV
Os10g0484800    NTVRTSLLFLGCSGGAAALRTAKDIAENN-PGSRVLVVAAETTVLGFRPPSPDRPYDLVGA
PKSA            DVNRVMLYFLGCYGGVTGLRVAKDIAENN-PGSRVLLTTSETTILGFRPPNKARPYDLVGA
AhCHSL1         DVNRVMLYFLGCYGGVTGLRVAKDIAENN-PGSRVLVTTSETTILGFRPPNKARPYDLVGA
AtCHSL2         EVQRVMLYFLGCYGGLSGLRVAKDIAENN-PGSRVLLTTSETTVLGFRPPNKARPYNLVGA
NsCHSLK         DIGRVMLYFLGCYGGVTGLRVAKDIAENN-PGSRVLLTTSETTILGFRPPNKARPYDLVGA
PrCHSL          DVSRVMLYFLGCYGGVTGLRVAKDLAENN-PGSRVLLATSETTILGFRPPNPERPYDLVGA
PpASCL          DVNRVMLYMLGCYGGASGIRVAKDLAENN-PGSRVLLITSECTLIGYKSLSPDRPYDLVGA 240
```

```
MsCHS          YVKRYMMYQQGCFAGGTVLRLAKDLAENN-KGARVLVVCSEVTAVTFRGPSDTHLDSLVGQ 212
AhSTS          SVKRYMMYHQGCFAGGTVLRLAKDLAENN-KDARVLIVCSENTAVTFRGPNETDMDSLVGQ
Gh2PS          SVKRYMLYQQGCAAGGTVLRLAKDLAENN-KGSRVLIVCSEITAILFHGPNENHLDSLVAQ 217

TaCHSL         ALFGDGAGAAVVGTDP--TPLECPLFELHSALQRFLPGTEKTIDGRLTEEGIKFQLGRELP
HvCHSL         ALFGDGAGAAVVGADP--TAVERPLFELHSALQRFLPDTEKTIDGRLTEEGIKFQLGRELP
ZmCHSL         ALFGDGAGAAVIGTDP--APAERPLFELHSALQRFLPDTERTIEGRLTEEGIKFQLGRELP
PKSB           ALFGDGAGAMIIGSDPD-PICEKPLFELHTAIQNFLPETEKTIDGRLTEQGINFKLSRELP
AhCHSL2        ALFGDGAGAMIIGSDPD-PVCEKPLFELHTAIQNFLPDTEKTIDGRLTEQGINFKLARELP
VvCHSL         ALFGDGAGAMIIGSDPI-PSTERPLFELHTAIQNFLPDTEKTIDGRLTEEGISFKLARELP
SlCHSL         ALFGDGAGAMIIGSDP--NSSENPLFELHTAIQHFLPDTEKIIDGRLTEEGISFTLDRALP
Os10g0484800   ALFGDGASAAIIGAGPI-AAEESPFLELQFSTQEFLPGTDKVIDGKITEEGINFKLGRDLP
PKSA           ALFGDGAAAVIIGADPR-EC-EAPFMELHYAVQQFLPGTQNVIEGRLTEEGINFKLGRDLP
AhCHSL1        ALFGDGAAAVIIGADPR-EC-EAPFMELHYAVQQFLPGTQNVIDGRLTEEGINFKLGRDLP
AtCHSL2        ALFGDGAAALIIGADPT-ES-ESPFMELHCAMQQFLPQTQGVIDGRLSEEGITFKLGRDLP
NsCHSLK        ALFGDGAAAVIIGTEPI-MGKESPFMELNFATQQFLPGTNNVIDGRLTEEGINFKLGRDLP
PrCHSL         ALFGDGAAAMVLGTDPRPEAGEQGFLELDWAVQQFLPDTHGTINGRLTEEGINFKLGRELP
PpASCL         ALFGDGAAAMIMGKDP-IPVLERAFFELDWAGQSFIPGTNKTIDGRLSEEGISFKLGRELP 300

MsCHS          ALFGDGAAALIVGSDPVPEI-EKPIFEMVWTAQTIAPDSEGAIDGHLREAGLTFHLLKDVP 272
AhSTS          ALFGDGAAAIIIGSDPVPEV-ENPIFEIVSTDQQLVPNSHGAIGGLLREVGLTFYLNKSVP
Gh2PS          ALFGDGAAALIVGSGPHLAV-ERPIFEIVSTDQTILPDTEKAMKLHLREGGLTFQLHRDVP 277

TaCHSL         HIIEAHVESFCQKLIKEHPAAAAAEGD--NMLTYDKMFWAVHPGGPAILTKMEGRLGLDGG
HvCHSL         HIIEAHVESFCQKLIKEHPGAAAAE-D--VPLTYDKMFWAVHPGGPAILTKMEGRLGLDGG
ZmCHSL         HIIEAHVEDFCQKLMKERQSGEDADGGGPEPMSYGDMFWAVHPGGPAILTKMEGRLGLGAD
PKSB           QIIEDNVENFCKKLIGKAG---LAHKN------YNQMFWAVHPGGPAILNRIEKRLNLSPE
AhCHSL2        QIIEDNVENFCKKLIGKAG---LAHKN------YNQMFWAVHPGGPAILNRMEKRLNLSPE
VvCHSL         QIIEDHIEAFCDKLIRNVG---FSDED------YNKIFWAVHPGGPAILNRMEKRLDLLPE
SlCHSL         QIIEDNIEAFCDKLMSSVG---LTSKD------YNDMFWAVHPGGPAILNRLEKRLDLSPD
Os10g0484800   EKIENRIEGFCRTLMDRVG-----------IKEFNDVFWAVHPGGPAILNRLEVCLELQPE
PKSA           QKIEENIEEFCKKLMGKAG-D--------ESMEFNDMFWAVHPGGPAILNRLETKLKLEKE
AhCHSL1        QKIEENIEEFCKKLMGKAGGD--------ESMEFNDMFWAVHPGGPAILNRLETKLKLEKE
AtCHSL2        QKIEDNVEEFCKKLVAKAG-S--------GALELNDLFWAVHPGGPAILSGLETKLKLKPE
NsCHSLK        EKIQDNIEEFCKKIIAKADLR--------E-AKYNDLFWAVHPGGPAILNRLENTLKLQSE
PrCHSL         QIIEDHIEGFCRKLMDKAG-----------VDDYNELFWGVHPGGPAILNRLEKKLSLGPE
PpASCL         KLIESNIQGFCDPILKR-----------AGGLKYNDIFWAVHPGGPAILNAVQKQLDLAPE 350

MsCHS          GIVSKNITKALVEAFEPLG-----------ISDYNSIFWIAHPGGPAILDQVEQKLALKPE 322
AhSTS          DIISQNINGALSKAFDPLG-----------ISDYNSIFWIAHLGGRAILDQVEQKVNLKPE
Gh2PS          LMVAKNIENAAEKALSPLG-----------ITDWNSVFWMVHPGGRAILDQVERKLNLKED 327
```

```
TaCHSL          KLRASRSALRDFGNASSNTIVYVLENMVE---ESRQR--AEAPEPE----------
HvCHSL          KLRASRSALRDFGNASSNTIVYVLENMVE---ESRRQRMTEAPVLENMVEESRRQRT
ZmCHSL          KLRASRCALRDFGNASSNTIVYVLENMVE---DTRRRR---------LL--------
PKSB            KLSPSRRALMDYGNASSNSIVYVLEYMLE---ESKKVRN-----------------
AhCHSL2         KLSPSRRALMDYGNASSNSIVYVLEYMLE---ESKKVRN-----------------
VvCHSL          KLNASRRALADYGNASSNTIVYVLEYMLE---ESSKTKR-----------------
SlCHSL          KLSASRRALTDYGNASSNTIVYVMEYMIE---EGLKRKN-----------------
Os10g0484800    KLKISRKALMNYGNVSSNTVFYVLEYLRD---ELKK--------------------
PKSA            KLESSRRALVDYGNVSSNTILYVMEYMRD---ELKKK-------------------
AhCHSL1         KLESSRRALVDYGNVSSNTILYVMEYMRE---ELKKK-------------------
AtCHSL2         KLECSRRALMDYGNVSSNTIFYIMDKVRD---ELEKK-------------------
NsCHSLK         KLDCSRRALMDYGNVSSNTIFYVMEYMRE---ELKNK-------------------
PrCHSL          KLYYSRQALADYGNASSNTIVYVLDAMRQ----LKGG-------------------
PpASCL          KLQTARQVLRDYGNISSSTCIYVLDYMRHQSLKLKEANDN----------------- 390

MsCHS           KMNATREVLSEYGNMSSACVLFILDEMRK---KSTQN------------------- 356
AhSTS           KMKATRDVLSNYGNMSSACVFFIMDLMRK---KSLET-------------------
Gh2PS           KLRASRHVLSEYGNLISACVLFIIDEVRK---RSMAE------------------- 361
```

```
TaCHSL          PEPEGADQECEWGLILAFGPGITFEGILARNLQARLGAN---
HvCHSL          TEPEM-EPECEWGLILAFGPGITFEGILARNLQARIAAN---
ZmCHSL          AADDG-GEDCEWGLILAFGPGITFEGILARNLQATARASAQL
PKSB            ----MNEEENEWGLILAFGPGVTFEGIIARNLDV--------
AhCHSL2         ----MNEEENEWGLILAFGPGVTFEGIIARNLDV--------
VvCHSL          ----QDQGDGEWGLILAFGPGITFEGILARNLTV--------
SlCHSL          ----GDKNDNDWGLILAFGPGL--------------------
Os10g0484800    -----GMIREEWGLILAFGPGITFEGMLVRGIN---------
PKSA            -----GDAAQEWGLGLAFGPGITFEGLLIRSLTSS-------
AhCHSL1         -----GDAAQEWGLGLAFGPGITFEGLLIRSL----------
AtCHSL2         -----GTEGEEWGLGLAFGPGITFEGFLMRNL----------
NsCHSLK         -----KNGGEEWGLALAFGPGITFEGILLRSL----------
PrCHSL          -----EKQSPEWGLILAFGPGITFEGILARSLV---------
PpASCL          -----VNTEPEWGLLLAFGPGVTIEGALLRNLC--------- 418

MsCHS           -GLKTTGEGLEWGVLFGFGPGLTIETVVLRSVAI-------- 389
AhSTS           -GLKTTGEGLDWGVLFGFGPGLTIETVVLRSMAI--------
Gh2PS           -GKSTTGEGLDCGVLFGFGPGMTVETVVLRSVRVTAAVANGN 402
```