Received Date : 26-Oct-2011
Revised Date : 10-Feb-2012
Accepted Date : 15-Feb-2012
Article type : Original Article

The hexanoyl-CoA precursor for cannabinoid biosynthesis is formed by an acyl-activating enzyme in *Cannabis sativa* trichomes

Jake M. Stout[1,2], Zakia Boubakir[1,2], Stephen J. Ambrose[1], Randy W. Purves[1] and Jonathan E. Page[1,2*]

[1]Plant Biotechnology Institute, National Research Council of Canada, 110 Gymnasium Place, Saskatoon, SK, S7N 0W9 Canada

[2]Department of Biology, University of Saskatchewan, 112 Science Place, Saskatoon, SK, S7N 5E2 Canada

* To whom correspondence should be addressed: Tel: 306-975-4187, Fax: 306-975-4839, Email: jon.page@nrc-cnrc.gc.ca

GenBank accession numbers for the nucleotide sequences reported in this study are: CsAAE1-11 (JN717233-JN717243) and CsAAE12-15 (JN859185-JN859188).

E-mail address of authors:

Jake Stout, jake.stout@nrc-cnrc.gc.ca; Zakia Boubakir, z.boubakir@tu-braunschweig.de; Stephen Ambrose, steve.ambrose@nrc-cnrc.gc.ca; Randy Purves, randy.purves@nrc-cnrc.gc.ca; Jonathan Page: jon.page@nrc-cnrc.gc.ca

Running title:

A cytoplasmic acyl-activating enzyme involved in cannabinoid biosynthesis

Key words:

Cannabis sativa, marijuana, cannabinoid, acyl-CoA, acyl-activating enzyme, hexanoate

SUMMARY

The psychoactive and analgesic cannabinoids (e.g. Δ^9-tetrahydrocannabinol, THC) in *Cannabis sativa* are formed from the short-chain fatty acyl-CoA precursor hexanoyl-CoA. Cannabinoids are synthesized in glandular trichomes present mainly on female flowers. We quantified hexanoyl-CoA using LC-MS/MS and found levels of 15.5 pmol g^{-1} fresh weight in female hemp flowers with lower amounts in leaves, stems and roots. This pattern parallels the accumulation of the end-product cannabinoid, cannabidiolic acid (CBDA). To search for the acyl-activating enzyme (AAE) that synthesizes hexanoyl-CoA from hexanoate, we analyzed the transcriptome of isolated glandular trichomes. We identified 11 unigenes encoding putative AAEs, including *CsAAE1* which shows high-transcript abundance in glandular trichomes. *In vitro* assays show that recombinant CsAAE1 activates hexanoate and other short- and medium-chain fatty acids. This activity and the trichome-specific expression of *CsAAE1* suggests that it is the hexanoyl-CoA synthetase that supplies

the cannabinoid pathway. CsAAE3 encodes a peroxisomal enzyme that activates a variety of fatty acid substrates including hexanoate. Although phylogenetic analysis shows that CsAAE1 groups with peroxisomal AAEs, it lacks a peroxisome targeting sequence 1 (PTS1) and localizes to the cytoplasm. We suggest that CsAAE1 may have been recruited to the cannabinoid pathway through the loss of its PTS1, thereby redirecting it to the cytoplasm. To probe the origin of hexanoate, we analyzed the trichome EST dataset for enzymes of fatty acid metabolism. The high abundance of transcripts encoding desaturases and a lipoxygenase suggests that hexanoate may be formed through a pathway involving the oxygenation and breakdown of unsaturated fatty acids.

INTRODUCTION

Cannabis sativa L. (hemp, marijuana; Cannabaceae) is well known for its content of cannabinoids, a group of more than 70 natural products with unique pharmacological properties (Elsohly and Slade 2005). Major cannabinoids include Δ^9-tetrahydrocannabinol (THC), the compound responsible for the psychoactive and therapeutic effects of marijuana consumption, and cannabidiol (CBD), which has neuroprotective properties (Gaoni and Mechoulam 1964; Mechoulam *et al.*, 2002). Cannabinoids are prenylated polyketides of mixed biosynthetic origin that are synthesized from fatty acid and isoprenoid precursors (**Figure 1**). Hexanoyl-CoA, derived from the short-chain fatty acid hexanoate, is used as a primer for a polyketide synthase (PKS) enzyme that forms olivetolic acid (OA) (Dewick 2009). Although OA formation has been investigated (Raharjo *et al.*, 2004; Marks *et al.*, 2009) and a candidate type III PKS (olivetol synthase, OLS) identified (Taura *et al.*, 2009), this step is unclear. OA is geranylated to yield cannabigerolic acid (CBGA) (Fellermeier

and Zenk 1998). The oxidocyclase tetrahydrocannabinolic acid synthase (THCAS) converts CBGA to THCA while cannabidiolic acid synthase (CBDAS) forms CBDA (Sirikantaramas *et al.*, 2004; Taura *et al.*, 2007). THCA is the major cannabinoid in drug strains of cannabis while CBDA predominates in hemp forms grown for fibre or seed. The cannabinoid acids occur *in planta* and undergo decarboxylation to their neutral forms upon heating or pyrolysis.

Acyl-CoA thioesters such as hexanoyl-CoA are formed by members of the acyl-activating enzyme (AAE) superfamily that activate carboxylic acids through an adenylate intermediate (Schmelz and Naismith 2009). In plants the substrates of AAEs include phenylpropanoids, fatty acids and jasmonate precursors. 4-Coumarate:CoA ligase (4CL), an enzyme involved in phenylpropanoid metabolism, and the long-chain acyl CoA synthetases, are perhaps the best characterized of the plant AAEs (Shockey *et al.*, 2002; Hu *et al.*, 2010). The genomic organization and biochemical diversity of AAEs has been recently reviewed (de Souza *et al.*, 2008; Shockey and Browse 2011).

The primary site of cannabinoid biosynthesis is glandular trichomes that form on female flowers (Lanyon *et al.*, 1981). Glandular trichomes are also found at lower densities on the male flowers, where a row of trichomes is located on the inner surfaces of anthers (Dayanandan and Kaufman, 1976). Analysis of the transcriptome of glandular trichomes from mint, hop and other plants through EST sequencing has been invaluable for discovering enzymes involved in specialized metabolic pathways (Lange *et al.*, 2000;

Nagel *et al.*, 2008; Wang *et al.*, 2008; Dai *et al.*, 2010). A recent EST analysis of cannabis trichomes has provided data on their metabolic functions (Marks *et al.*, 2009).

We sought to clarify the formation of hexanoyl-CoA as precursor for cannabinoid biosynthesis. LC-MS/MS analysis established that hexanoyl-CoA is present in cannabis flowers. To identify the enzyme(s) responsible for hexanoyl-CoA synthesis, we used transcriptome analysis of a trichome-specific cDNA library from female flowers to identify 11 putative *AAE* genes. Here we report that two encode hexanoyl-CoA synthetases and provide evidence that CsAAE1 is a cytoplasmic enzyme that functions in cannabinoid biosynthesis. CsAAE3 encodes a peroxisomal acyl-CoA synthetase that activates a variety of substrates including hexanoate.

RESULTS

Hexanoyl-CoA and CBDA are present in female cannabis flowers

Hexanoyl-CoA has not previously been identified in cannabis and we sought to measure this metabolite in tissues actively synthesizing cannabinoids. We used plants of the hemp cultivar 'Finola' which has a cannabinoid profile typical of hemp varieties (high CBDA/low THCA) (**Figure S1**).The analysis was initially performed using trichome cells, but measurement of acyl-CoAs was not reproducible, possibly because of losses during trichome isolation. Instead we found that hexanoyl-CoA was readily detectable when whole flowers with trichomes were analyzed (**Figure 2**). After adjusting for recoveries using the internal standards, we quantified the hexanoyl-CoA pools in female flowers, leaves, stems and roots. To compare these pool sizes to those of the final pathway product, we also quantified the content of the cannabinoid pathway end-product CBDA in the same tissues.

As shown in **Table 1**, the levels of hexanoyl-CoA and CBDA are highest in female flowers and substantially lower in the other tissues analyzed.

Cannabinoid pathway transcripts are abundant in cannabis trichomes

We analyzed the transcriptome of cannabis trichome cells for genes encoding enzymes of cannabinoid and terpenoid biosynthesis, with a focus on *AAE* transcripts. The trichomes were obtained exclusively from female flowers, which are larger and have more trichomes than male flowers. A trichome-specific cDNA library from a marijuana strain was sequenced to produce 9157 ESTs that assembled into 4113 unique sequences (1227 contigs, 2886 singletons). Unigenes were annotated by blastx comparison to the UniProt database. Transcripts encoding cannabinoid biosynthetic enzymes were abundant in the EST dataset: *OLS* (Taura *et al.*, 2009) ranked first with 340 contig members and *THCAS* (Sirikantaramas *et al.*, 2004) ranked eleventh with 48 contig members (**Table 2**). Genes for enzymes of the 2-*C*-methyl-D-erythritol 4-phosphate (MEP) pathway, which produces the geranyl side-chain of the cannabinoids, were highly expressed (Fellermeier *et al.*, 2001; Phillips *et al.*, 2008). We also observed that transcripts representing several desaturases and enzymes of fatty acid degradation, specifically a lipoxygenase and an acetoacetyl-CoA thiolase, were prominently represented in the trichome dataset.

Cannabis AAE genes (*CsAAEs*) were identified by utilizing Arabidopsis AAE sequences to query the assembled cannabis ESTs using tblastn. Eleven unique *CsAAE* contigs and singletons were identified and named according to their abundance in the cDNA library (**Table 3**). We considered two to be strong candidates to encode a hexanoyl-CoA

synthetase. *CsAAE1* was the most highly expressed *AAE* based on transcript abundance (42 ESTs), and the sixteenth most abundant transcript in the EST dataset. *CsAAE1* also had the highest EST count of the four AAE genes present in the cannabis trichome EST dataset reported by Marks *et al.* (2009). The second candidate, *CsAAE3*, was identified due to its homology to an Arabidopsis *4CL-like* gene (At4g05160) that activates fatty acids including hexanoate (Schneider *et al.*, 2005). Of the nine other AAEs, only CsAAE2, which was not full-length, showed high EST counts. This protein was most similar to the Arabidopsis LACS4 AAE (At4g23850), which activates long-chain fatty acids (Shockey *et al.*, 2003).

CsAAE1 and CsAAE3 encode hexanoyl-CoA synthetases

Of the eight full-length *AAE* cDNAs, *CsAAE1*, *CsAAE3*, *CsAAE6*, *CsAAE8* and *CsAAE10* yielded soluble protein when expressed in *E. coli* (**Figure S2**). *CsAAE5*, *CsAAE7* and *CsAAE9* formed inclusion bodies are were not analyzed further. The hexanoyl-CoA synthetase activity of the purified recombinant proteins was tested with hexanoate, CoA, ATP and $MgCl_2$ (**Figure 3**). CsAAE1 and CsAAE3 formed hexanoyl-CoA, whereas the other three enzymes were inactive. Both CsAAE1 and CsAAE3 exhibited temperature and pH optima of 40°C and pH 9, respectively (**Figure S3**). CsAAE1 activity was highest with Mg^{2+}. Although CsAAE3 preferred Co^{2+}, Mg^{2+} was used for further assays.

To test the substrates that CsAAE1 and CsAAE3 can activate, we assayed a range of carboxylic acids using luciferase-based assay system (Schneider *et al.*, 2005). CsAAE1 utilized hexanoate, heptanoate, octanoate and nonanoate, with a slight preference for hexanoate (**Figure 4**). In contrast, CsAAE3 accepted short-, medium-, and long-chain fatty

acids. To further this analysis, the kinetic properties of CsAAE1 and CsAAE3 were measured for CoA and representative short (butanoate), medium (hexanoate and decanoate), and long-chain (palmitate) fatty acids (**Table 4, Figure S4**). High CoA concentrations inhibited CsAAE1. Using a non-linear regression substrate inhibition model, we estimate that the K_i of CsAEE1 5.101 ± 1.8 mM. CoA did not inhibit CsAAE3 at the concentrations tested. Decanoic acid inhibited CsAAE3 (K_i = 120.8 ± 47.9 μM) and slightly inhibited CsAAE1 in concentrations above 4 mM (K_i not measured).

CsAAE1 transcript is abundant in glandular trichomes

CsAAE1 and *CsAAE3* transcripts were measured in different organs of 'Finola' plants. Total RNA was isolated from roots, stems, leaves, female flowers (before and after trichomes removal), male flowers and trichome cells. *CsAAE1* transcript levels were high in trichome cells and low in the other plant parts (**Figure 5**). The trichome-specific expression was further supported by the finding that removal of trichomes from the female flowers reduced *CsAAE1* transcript levels. The expression pattern of *CsAAE1*, including the reduced expression in female flowers with trichomes removed (FF-) compared to intact female flowers (FF+), is similar to that of the known cannabinoid pathway gene *CBDAS*. *CsAAE3*, which was present in all tissues (**Figure 5**), does not show the same expression pattern as *CsAAE1* or *CBDAS*. *CsAAE1* transcripts were more abundant in trichome cells than *CsAAE3*, a finding that is also reflected in their EST counts (**Table 3**). The residual *CsAAE1* and *CBDAS* expression in FF- samples may reflect in the incomplete removal of trichome secretory cells from these tissues, or the possibility that some cannabinoid biosynthesis occurs in internal floral tissues.

CsAAE1 and CsAAE3 are localized to different subcellular compartments

Many plant AAEs are located in the peroxisome where they function to activate substrates for β-oxidation (Shockey *et al.*, 2003; de Souza *et al.*, 2008). Peroxisomal proteins either contain a C-terminal peroxisome targeting sequence 1 (PTS1) of twelve amino acids ending with SKL (or variants thereof), or an N-terminal peroxisome targeting sequence 2 (PTS2) sequence with a RLx5HL motif (Reumann 2004). CsAAE1 or CsAAE3 were analyzed for PTSs using a BLOCKS-based algorithm (www.peroxisomedb.org). The C-terminal sequence of CsAAE1 is not predicted to have a functional PTS1 signal (e-value 0.32), whereas CsAAE3 contains a non-canonical PTS1 sequence (e-value 0.04). Neither protein contains a PTS2 sequence.

We sought to verify these predictions by co-localizing fluorescent protein fusions of the CsAAE1 and CsAAE3 with organelle-specific marker proteins using transient expression in *Nicotiana benthamiana* leaves (Sparkes *et al.*, 2006). To avoid disrupting the C-terminal PTS1 signal, CsAAE1 and CsAAE3 were expressed as N-terminal yellow fluorescent protein (YFP) fusions. When co-infiltrated leaves were examined using confocal microscopy, YFP:CsAAE3 was found in discrete compartments that co-localized with the PX-CK peroxisome cyan fluorescent protein (CFP) marker protein (**Figure 6**). In contrast, YFP:CsAAE1 localized to the cytosol. We compared the localization of CsAAE1 with the putative cannabinoid pathway enzyme OLS, which does not contain targeting sequences and is therefore likely to be in the cytosol. The co-infiltration of YFP:CsAAE1 and

OLS:CFP resulted in overlapping fluorescence signals for YFP and CFP, indicating that these proteins are cytosolic (**Figure 6**).

Phylogenic relationships of cannabis AAEs

We examined the evolutionary relationships between the CsAAEs and those from other plants. We first generated a maximum likelihood phylogenetic tree using cannabis and Arabidopsis AAE protein sequences (Shockey *et al.*, 2003; Shockey and Browse 2011) (**Figure S5**). CsAAE1 is a member of clade II, which includes AtAAE17 (At5g23050), AtAAE18 (At1g55320) and a plastidic acetyl-CoA synthetase (ACS, At5g36880). CsAAE3 grouped with the clade V proteins, which includes enzymes that accept a range of fatty acids, hormone precursors and phenylpropanoids as substrates (de Souza *et al.*, 2008; Kienow *et al.*, 2008; Koo *et al.*, 2006). All clade V AAEs possess canonical PTS1 signals and are predicted to be located in the peroxisome.

To gain further insight into the evolution of CsAAE1, we expanded the analysis of AAE clade II to include homologs from *Chlamydomonas*, basal plants and higher eudicots. Using data from the recent sequencing of the cannabis genome (van Bakel *et al.*, 2011), we identified three additional clade II AAEs (CsAAE12, 13 and 14). The proteins in the expanded clade II group into three distinct subclades (**Figure 7A**). We designated the group that included CsAAE1 and AtAAE17 as subclade IIa, the group that included AtAAE18 as subclade IIb, and the group that contained ACS and homologs as subclade IIc. To investigate if peroxisomal localization is a hallmark of subclades IIa/b, the C-terminal twelve amino acids of each protein in the clade IIa/b phylogeny were aligned using Muscle,

and the presence of a functional PTS1 signal was assessed using a BLOCKS-based algorithm (www.peroxisomedb.org) to which experimentally-validated plant PTS1 sequences were added (Reumann 2004; Lingner *et al.*, 2011). In total, 29 out of 30 were predicted to have a functional PTS1 (e-values ≤0.25), indicating that all but one of the clade IIa/b AAEs are targeted to the peroxisome (**Figure 7B**). CsAAE1 is the only clade II AAE that lacks a PTS1.

DISCUSSION

Role of hexanoyl-CoA in cannabinoid biosynthesis

Cannabinoids are a rare example of a plant polyketide biosynthesized from hexanoate or hexanoyl-CoA precursors. Many plants and microbes use fatty acyl-CoAs as polyketide primers (e.g. alkylresorcinols in *Sorghum* (Cook *et al.*, 2010)), yet there are only a few examples of hexanoate/hexanoyl-CoA-derived polyketides in plants and fungi. Primin, the allergenic principle of *Primula obconica*, is postulated to be formed from hexanoyl-CoA (Horper 1996). Outside of the plant kingdom, *Dictyostelium discoideum* synthesizes the DIF-1 polyketide from hexanoyl-acyl carrier protein (ACP) thioester (Austin *et al.*, 2006; Ghosh *et al.*, 2008). The well-known fungal toxins of the aflatoxin/sterigmatocystin class are formed from hexanoate (Townsend 1984; Yabe and Nakajima 2004) as are the antitumor benastatins in *Streptomyces* (Xu *et al.*, 2009).

The detection of hexanoyl-CoA in cannabis flowers demonstrates that this compound is present in the tissues where cannabinoid biosynthesis occurs (**Table 1**). The concentration

of hexanoyl-CoA was the highest in female flowers, which also contained the highest concentration of CBDA. CBDA was also detected in leaves, stems and roots, albeit at concentrations that were orders of magnitude less compared to flowers.

The small size of the pool, measured to be 15.5 pmol g^{-1} fresh weight, is perhaps not surprising considering that this compound is a pathway intermediate. This analysis may also be complicated by the possible association of hexanoyl-CoA with acyl-CoA binding protein (Xiao and Chye 2011). To our knowledge, there exist only a few reports that describe the quantification of short- and medium-chain acyl-CoA thioesters in plants. Perera *et al.* (2009) used LC-MS/MS to quantify the butyryl-CoA/isobutyryl-CoA content of Arabidopsis siliques and leaves to be 380 pmol g^{-1} and 160 pmol g^{-1} fresh weight, respectively. These pools are approximately an order of magnitude greater than what we observed in cannabis flowers.

CsAAE1 is a cannabinoid biosynthetic gene that neofunctionalized by different subcellular localization

Through a sequence-based analysis of the trichome EST dataset and biochemical assay of five AAEs, we identified two that possess hexanoyl-CoA synthetase activity. Our data provides evidence that CsAAE1 is the enzyme involved in cannabinoid biosynthesis. *CsAAE1* was the most abundant transcript in the EST catalog, and qRT-PCR data shows that its expression is over 100-fold higher in trichome cells compared to other tissues (**Figure 5**). Furthermore, we have localized CsAAE1 to the cytosol, which is the same compartment where the putative cannabinoid enzyme OLS is localized. The substrate

preference of CsAAE1 provides additional evidence for its role in cannabinoid biosynthesis since it shows more specificity for hexanoate and other short-chain fatty acyl CoAs than CsAAE3 (**Figure 4**). Unfortunately, there have been no reports of successful genetic transformation of cannabis, and it was not possible to probe the function of CsAAE1 or CsAAE3 by using transgenic approaches such as RNAi. Our attempts at virus-induced gene silencing (VIGS) using tobacco rattle virus were also not successful (data not shown).

An argument against the involvement of CsAAE1 in cannabinoid biosynthesis is its high K_m value (3.7 mM) for hexanoate. K_m values from plant AAEs tend to be in the low μM range (**Table S1**) but most of these values are from clade IV and V AAEs, and it is not clear if these values are consistent for members of other AAE clades. The Arabidopsis malonyl-CoA synthetase (AtAAE13; clade VII) has a K_m for malonate of 529 μM, which is approximately seven-fold lower than that of CsAAE1 for hexanoate (Chen *et al.*, 2011). Analysis of the kinetic properties of CsAAE1 showed that this enzyme could only utilize hexanoate and, to a lesser extent, decanoate as substrate, with hexanoate acid yielding the highest k_{cat} value (**Table 4**). These data are corroborated with those from the luciferase-based ATP depletion assay (**Figure 4**). In contrast to CsAAE1, CsAAE3 accepted a broader range of substrates (**Figure 4**) and showed reasonable kinetic properties for hexanoate, decanoate and palmitate (**Table 4**). Interestingly, CsAAE3 was inhibited by decanoate but not hexanoate or palmitate. A similar inhibitory effect of C8-C12 fatty acids has been observed for CsAAE3's close Arabidopsis homolog (At4g05160) (Schneider *et al.*, 2005).

One route to overcoming catalytic inefficiency and low turnover of secondary metabolic enzymes is to increase expression levels (reviewed in Pichersky and Gang (2000)). Although we did not quantify the CsAAE1 protein, the high transcript levels indicated by ESTs counts (**Table 2**) and qRT-PCR (**Figure 5**) shows that it is highly expressed.

CsAAE1 lacks a PTS1 (**Figure 7B**) and its recruitment to the cannabinoid pathway may have resulted from altered subcellular localization. Given that all other clade II AAEs are localized in the peroxisome or plastid (i.e. clade IIc), it is possible that this neofunctionalization is a recent evolutionary event specific to cannabis. *CsAAE1* may have arisen through a tandem duplication event and has undergone subsequent divergence from its ancestral function through a change in subcellular localization. The identification of a second clade IIa cannabis AAE (*CsAAE12*) through our cannabis genome sequencing effort (van Bakel *et al.*, 2011) supports this hypothesis. *CsAAE12* is adjacent to *CsAAE1* in the genome, encodes a protein that has 77% amino acid identity to CsAAE1 and possesses a functional PTS1 signal (**Figure 7B**). Alternatively, *CsAAE1* may have evolved through allelic drift away from the ancestral clade IIa function.

CsAAE3 is a peroxisomally-targeted enzyme that is likely involved in β-oxidation

We found that CsAAE3 is more efficient than CsAAE1 at synthesizing hexanoyl-CoA. However, CsAAE3 is localized to the peroxisome (**Figure 6**) and it is not clear how hexanoyl-CoA formed in this compartment could be exported to the cytoplasm where the polyketide synthesis phase of the cannabinoid pathway is located. CsAAE3 accepts a very

broad range of substrates, indicating that it is a more generalized acyl-CoA synthetase that may function in peroxisomal β-oxidation.

Origin of hexanoate

Although the incorporation of ^{13}C-label into cannabinoids supports the hypothesis that hexanoate is formed from fatty acid precursors (Fellermeier *et al.*, 2001), the origin of hexanoate for activation by CsAAE1 has not been clarified. There are several possible metabolic routes to short- and medium-chain fatty acids in plants: *de novo* biosynthesis, elongation of α-keto acids, and fatty acid degradation. Marks *et al.* (2009) concluded that hexanoate was likely formed from the *de novo* pathway based on the high expression of ACP and 3-keto-ACP reductase in trichomes compared to leaves. The *de novo* route would require an acyl-ACP thioesterase to terminate fatty acid biosynthesis at a C6 fatty acid. Although C8- and C10-specific thioesterases are found in plants that accumulate medium-chain fatty acids (e.g. *Cuphea*; Dehesh *et al.*, 1996), there are no C6-specific thioesterases reported. We searched the cannabis trichome transcriptome and found three contigs that were similar to acyl-ACP thioesterases. Represented by only eight, three and two ESTs, they were not as prominent as the cannabinoid and terpenoid pathway transcripts. We were unable to detect any thioesterase-like sequences in the EST dataset reported by Marks *et al.* (2009). A second possible route to hexanoate is the degradation of branched chain amino acids and elongation of the resulting α-keto acids (α-KA) (Kroumova *et al.*, 1994; Kroumova and Wagner 2003). We considered this route unlikely given that ^{13}C feeding provided no evidence to support this route (Fellermeier *et al.*, 2001) and the lack of high expression for transcripts for branched-chain amino acid metabolism. A third option is the

formation of C18 unsaturated fatty acids by the action of desaturases, and their cleavage by lipoxygenase and hydroperoxide lyase to yield C6 and C12 products. Volatile C6-aldehydes in soybean, apple and tomato are formed by a lipoxygenase pathway (Matoba *et al.*, 1985; Gray and Prestage 1999; Rowan *et al.*, 1999). The high expression of three desaturases (62, 53 and 22 ESTs) and a lipoxygenase (25 ESTs) provides evidence that the fatty acid degradation pathway maybe active in trichomes, although hydroperoxide lyase was represented by three ESTs.

Metabolic engineering of cannabinoid biosynthesis

Our discovery of AAEs that catalyze the biosynthesis of hexanoyl-CoA and other acyl-CoA esters may further the breeding of low-cannabinoid cannabis varieties. To be permitted to be grown legally in Canada and EU, hemp cultivars must contain less than 0.3% of THCA. While breeding has led to low-THCA hemp, reducing CsAAE1 activity through targeted mutagenesis may lead to further decreases and also reduce other non-psychoactive cannabinoids such as CBDA.

Cannabinoids have received increased attention as drugs to treat a number of diseases and conditions (Mechoulam 2005; Izzo *et al.*, 2009). The discovery of hexanoyl-CoA synthetases, together with the identification of other genes encoding the cannabinoid biosynthetic pathway (**Figure 1**), allows for the reconstruction of the pathway in heterologous systems using a synthetic biology approach. Such a platform would enable the discovery of genes encoding enzymes that catalyze the formation of the less-abundant

cannabinoid derivatives, and may allow for scalable fermentation production of these compounds so that their efficacy can be tested.

MATERIALS AND METHODS

Plant cultivation and isolation of glandular trichomes

Female plants of a marijuana strain (~12% THCA by dry weight, low CBDA) were propagated from cuttings and grown in a secure controlled environment chamber under long day (18 hour) vegetative conditions. Flowering was induced by transfer to 12-hour day length. Female plants of the hemp cultivar 'Finola' (<0.3% THCA, high-CBDA) were grown from seed under 16-hour day conditions. Trichome cells were isolated from female flowers using a Beadbeater method adapted from (Gershenzon *et al.*, 1992). Yield was ca. 0.9 g from 60 g of flowers.

cDNA library and ESTs

Total RNA from trichomes from a high-THCA marijuana strain was isolated using a CTAB-based method (Meisel *et al.*, 2005) followed by clean-up using an RNeasy Midi (Qiagen) kit. Genomic DNA was removed by on-column digest with DNase I (Qiagen). mRNA was isolated from total RNA using Dynabeads Oligo(dT) (Invitrogen). mRNA (3.7 µg) was used as a template for cDNA library synthesis (Stratagene) to produce cDNAs with EcoR1 (5') and Xho1 (3') sites. cDNA was separated on a 1% agarose gel into >2.5 kb, 1.5-2.5 kb and <1.5 kb fractions and each ligated separately into pBluescript SK+ vector digested with EcoR1 and Xho1 before electroporation into *E. coli* DH10B T1 Phage Resistant Cells (Invitrogen). Bacterial cultures were used for TempliPhi amplifications (GE Healthcare). TempliPhi products

were sequenced with the T3 primer. BigDye reactions were resolved on an ABI 3730xl sequencer (Applied Biosystems). Sequences were processed using a high-performance cluster computer using the procedures described in (Nagel *et al.*, 2008). Sequences were annotated by blastx comparison to the UniProt database.

Analysis of hexanoyl-CoA and CBDA in cannabis flowers

A solution containing benzoyl-CoA and hexanoyl-CoA standards (1 µM in 95:5 water:methanol with 5 mM TEA and 3 mM acetic acid) was used to optimize chromatographic conditions. Benzoyl-CoA was used as the internal standard. Female 'Finola' plants were divided into flower, leaf, stem and root samples, which were then ground to homogeneity in liquid nitrogen. Approximately 400 mg of homogenized tissue was extracted with 5% trichloroacetic acid (TCAA) for 5 min at 4°C. The samples were centrifuged at 3000 rpm for 5 min and the supernatant purified using cation exchange (Waters MCX1CC Oasis cartridge): after activation with 1 mL methanol and 1 mL water, the TCAA supernatant was loaded, washed with 1 mL 0.1 N HCl and 1 mL 100% MeOH, and eluted with 1 mL 25 mM NH_4OH in 82% MeOH. The eluate was dried and the residue resuspended in 100 µL of the starting LC buffer. UPLC MS/MS analysis was performed using a Waters Acquity UPLC system with a diode array detector and a Quattro Ultima mass spectrometer in electrospray ionization (ESI) multiple reaction monitoring (MRM) positive ion mode. Chromatography was performed on a reverse phase UPLC column (2.1 × 50 mm, 1.7 µm, BEH C18, Waters) using 7.5 µL injections. Separations were achieved with a binary solvent system consisting of solvent A (water with 5 mM triethylamine and 3 mM acetic acid, pH 10) and solvent B (methanol and water (95:5) with 5 mM triethylamine

and 0.3 mM acetic acid, pH 11). The elution gradient (200 µL/min) was: 5.3% B for 1 min, 15% B at 6 min and 50% B at 10 min. The column was flushed and returned to starting conditions at 30 min. Consistent with previous work (Perera *et al.*, 2009), optimum MRM transitions in positive ionization mode resulted in a loss of 507 amu. Quantification was carried out by using a peak area ratio of the MRM transitions for hexanoyl-CoA (*m/z* 866 >359, 9.25 min) relative to benzoyl-CoA (*m/z* 872 >365, 8.45 min).

CBDA was measured in the same tissues used to quantify hexanoyl-CoA. Approximately 200 mg of homogenized tissue was extracted with 50% methanol at 65°C for 10 min. The extracts were filtered using a Spin-X column (Corning). HPLC analysis was performed using a Waters 2695 system coupled to a Waters 3100 single quadrupole mass detector and a 996 photodiode array detector. Separations were achieved on a reverse phase column (2.1 mm × 5 cm, 2.7 µm, Ascentis Express C18, Sigma-Aldrich) using a binary solvent system consisting of solvent A (water containing 10% acetonitrile, 0.01 % formic acid) and solvent B (99.09% acetonitrile, 0.01 % formic acid) with gradient elution (250 µL min^{-1}) of: 45% B from 0 to 8 min; 95% B from 8 min to 10.5 min, 45% B at 13 min with a hold to 20 min. Peaks were detected by absorbance at 270 nm. CBDA was quantified using a standard curve prepared from an authentic CBDA standard (40 to 9500 pmoles injected, r^2 1.0).

qRT-PCR of CsAAE1 and CsAAE3 expression

'Finola' plants were grown from seed until mid-flowering stage. Roots, stems, leaves, female flowers (with trichomes and after trichome isolation using the Beadbeater), trichome cells and male flowers were sampled from three plants. Total RNA was isolated as

described above. RNA had an Abs_{260}:Abs_{280} of >1.9 and showed distinct ribosomal bands on denaturing gel. First-strand cDNA were synthesized using 0.5 µg RNA with a QuantiTect cDNA Synthesis kit (Qiagen). Each 20 µL cDNA sample was diluted 1:4 with water, and 1 µL used as a PCR template. Gene-specific primers were designed to produce amplicons of 90-200 bp (**Table S2**). PCR reactions (20 µL) were performed in 96-well plates using a SYBR Green based assay (QuantiFast SYBR Green kit, Qiagen) with a StepOne Plus instrument (Applied Biosystems). The cycling parameters used were 95°C for 5 min followed by 40 cycles of 95°C for 10 s, 60°C for 30 s, and a standard dissociation protocol (95°C 15 s, 60°C for 1 min, 60-95°C in 0.3°C increments for 15 s). Experiments were performed using cDNAs from three plants with two technical replicates. *Actin*, which we found had stable expression in all tissues tested, was used as a reference gene. The efficiencies for all primer pairs were 90-110% as calculated using the standard curve method. C_t values were calculated using StepOne Software (Applied Biosystems). The $2^{-\Delta\Delta Ct}$ method was used for relative gene expression analysis.

Recombinant protein expression, purification and assay

CsAAE ORFs were amplified using gene-specific primers (**Table S2**) using full-length cDNAs or trichome cDNA as template. The PCR products were cloned into pCR8/GW/TOPO (Invitrogen) and recombined into the pHIS8/GW using LR recombinase (Invitrogen). Plasmids were transformed into Rosetta 2 (DE3) cells (Merck) and protein expressed in 500 mL LB medium with 0.2 µM IPTG. *CsAAE1* was expressed at 12°C for 24 h, whereas *CsAAE3* was expressed at 37°C for 16 h. Cells were harvested by centrifugation and resuspended in 10 mL His-tag lysis buffer (50 mM Tris-HCl pH 7, 500 mM NaCl, 2.5

mM imidazole, 10% v/v glycerol, 10 mM β-mercaptoethanol, 1% v/v Tween 20, 750 µg/mL lysozyme), incubated on ice for 1 h, and lysed by sonication. After centrifugation for 20 min at 12,000 g at 4ºC, the supernatent was added to 160 µL of Talon resin (Clontech) that had been washed with His-tag wash buffer (HWB; 50 mM Tris-HCl pH 7, 500 mM NaCl, 2.5 mM imidazole, 10% v/v glycerol, 10 mM β-mercaptoethanol). The samples were incubated with rocking at 4ºC and the resin pelleted by centrifugation. The resin was resuspended in HWB, washed at 4°C, and recovered by centrifugation. The wash step was repeated twice and the samples were loaded onto a chromatography column. After washing with 10 mL of HWB, his-tagged proteins were eluted with 2.5 mL of His-tag elution buffer (50 mM Tris-HCl pH 7, 500 mM NaCl, 250 mM imidazole, 10% v/v glycerol, 10 mM β-mercaptoethanol). The eluates were exchanged into 50 mM HEPES pH 9, 10% v/v glycerol, 2 mM $MgCl_2$, 2 mM dithiothreitol using PD10 columns (Amersham Biosciences). The purity of the isolated proteins was verified by SDS-PAGE (**Figure S1**), and the protein concentration was determined by Bradford assay.

Hexanoyl-CoA synthetase assays were performed in a 20 µL reaction containing 0.1 µg recombinant AAE, 50 mM HEPES pH 9, 8 mM ATP, 10 mM $MgCl_2$, 0.5 mM CoA and 4 mM sodium hexanoate for 10 min at 40°C. Reactions were terminated with 2 µL of 1 N HCl and stored on ice until analysis. After centrifugation at 15,000 *g* for 5 min at 4°C, the samples were diluted 1:100 in water and analyzed by UPLC-MS/MS as described above.

The kinetic properties of CsAAE1 and CsAAE3 with CoA were tested in assays containing 50 mM HEPES pH 9.0, 10 mM sodium hexanoate, 8 mM ATP, 10 mM $MgCl_2$, 0.02-2 mM

CoA, and 290 ng of CsAAE1 or 10 ng of CsAAE3. Detection of hexanoyl-CoA was performed by analyzing 5 µL of the reaction mixtures by HPLC-MS using a Waters 2695 system coupled to a Waters 3100 single quadrupole mass detector. The isocratic LC conditions consisted of 87% 2 mM ammonium formate pH 4.6 and 13% acetonitrile, with a flow rate of 250 µL/min. Compounds were separated on a reverse phase column (2.1 mm × 5 cm, 2.7 µm, Ascentis Express C18, Sigma-Aldrich). Hexanoyl-CoA was detected in positive mode by scanning for m/z of 866 Da (CV = 45 V, span = 0.2 Da, dwell time = 0.01 s). Curve fitting of the data was performed with Prism 5.0 (GraphPad) using the non-linear regression with substrate inhibition model for CsAAE1 and the Michaelis-Menten model for CsAAE3. Kinetics were measured in assays containing 50 mM HEPES pH 9.0, 8 mM ATP, 10 mM $MgCl_2$, and various concentrations of fatty acid. [CoA] was 0.5 mM for CsAAE1 and for 2 mM CsAAE3. Compounds were separated as above, except that isocratic conditions were 93% ammonium formate for butyryl-CoA, 72% for decanoyl-CoA and 45% for palmitoyl-CoA analysis. Masses scanned were 838 Da (butyryl-CoA), 922 Da (decanoyl-CoA), and 1006.5 (palmitoyl-CoA). Reaction products were quantified using standard curves of authentic CoA-thioester standards (r^2 of 1.0, butyryl-CoA; 0.99, hexanoyl-CoA; 1.0, decanoyl-CoA; 0.99, palmitoyl-CoA).

Substrate preference assay

Assays conditions were identical to those used in (Schneider *et al.*, 2005). To each well of a 96-well plate, 90 µL of 100 mM Tris pH 7.8, 1 mM $MgCl_2$, 2.3 µg luciferin and 0.5 µg of luciferase were added. After shaking for 2 s the luminescence was measured with a 1420 Multilabel counter (PerkinElmer) for 15 s without a filter in place.

Subcellular localization of CsAAE1 and CsAAE3

YFP:CsAAE1 and YFP:CsAAE3 fusions were constructed by recombination into pEARLYGATE104 (Earley *et al.*, 2006) using LR recombinase (Invitrogen). To generate an OLS:CFP construct, *OLS* lacking a stop codon was cloned into pCR8/GW/TOPO before recombination into pEARLYGATE102 using LR recombinase. The peroxisome marker PX-CK (Nelson *et al.*, 2007) was from ABRC (www.arabidopsis.org). Plasmids were transformed into *Agrobacterium tumefaciens* GV3101 by electroporation and selected on LB plates containing 10 µg/mL rifampacin and 50 µg/mL kanamycin. Leaves of two-week old *Nicotiana benthamiana* plants were infiltrated with the Agrobacterium solution at an OD_{600} of 0.02 (Sparkes *et al.*, 2006). Two days post-infiltration, leaf epidermal cells were visualized using a Zeiss LSM510 confocal microscope. CFP was visualized with excitation 458 nm and image collection with a 475-525 nm bandpass filter; YFP at 514 nm with a 530-600 nm bandpass filter. Images were collected and analyzed using the Zeiss LSM software package.

Phylogenetic Analyses

Arabidopsis, *Physcomitrella* and *Populus* AAE sequences were those previously identified (Shockey and Browse 2011). Additional clade II homologs were retrieved by using blastp to compare the Arabidopsis clade II sequences against the Phytozome database (www.phytozome.net). Phylogenetic analyses were performed with the Mega v5.05 (Tamura *et al.*, 2011) using the Maximum likelihood algorithm with distances computed using 1000 bootstrap replicates.

Acknowledgements

Funding was provided by the Natural Sciences and Engineering Research Council of Canada, the National Research Council of Canada, Genome Canada, Genome Prairie and the Government of Saskatchewan. We thank J. Noel (Salk Institute) for pHIS8/GW, P. Covello for substrates, S. Qiu for cDNA library construction, R. Taschuk for cloning OLS, S. Polvi for growing hemp plants, D. Cram for bioinformatics, and M. Smith and J. Zou for comments on the manuscript.

Supporting Information

Supplemental Figure 1. HPLC chromatogram of hemp 'Finola' flower extract. Detection at 270 nm.

Supplemental Figure 2. SDS-PAGE analysis of purified recombinant AAEs.

Supplemental Figure 3. Biochemical properties of CsAAE1 and CsAA3.

Supplemental Figure 4. Kinetic plots of CsAAE1 and CsAAE3.

Supplemental Figure 5. Maximum likelihood inference of the phylogenetic relationship of cannabis AAEs and the Arabidopsis AAE superfamily. Amino acid sequences were aligned using Muscle and analyzed with Mega5. Branch point bootstrap values were calculated with 1000 replicates, and values less than 60 are not shown. Cannabis AAEs are highlighted in green.

Supplemental Table 1. *K*m values of acyl-activating enzymes reported in the literature.

Supplemental Table 2. Primers sequences used in this study.

Supplementary Table 3. Protein sequences used for phylogenetic analysis.

REFERENCES

Austin, M.B., Saito, T., Bowman, M.E., Haydock, S., Kato, A., Moore, B.S., Kay, R.R., and Noel, J.P. (2006). Biosynthesis of *Dictyostelium discoideum* differentiation-inducing factor by a hybrid type I fatty acid–type III polyketide synthase. *Nature Chem. Biol.* **2**, 494-502.

van Bakel, H., Stout, J.M., Cote, A.G., Tallon, C., Sharpe, A.G., Hughes, T.R. and Page, J.E. (2011). The draft genome and transcriptome of *Cannabis sativa*. *Genome Biol.* **12**, R102 doi:10.1186/gb-2011-12-10-r102.

Chen, H., Kim, H.U., Weng, H. and Browse, J. (2011). Malonyl-CoA synthetase, encoded by *ACYL ACTIVATING ENZYME13*, is essential for growth and development of *Arabidopsis*. *Plant Cell* **23**, 2247-2262.

Cook, D., Rimando, A.M., Clemente, T.E., Schroder, J., Dayan, F.E., Nanayakkara, N., Pan, Z., Noonan, B.P., Fishbein, M., Abe, I., Duke, S.O. and Baerson, S.R. (2010). Alkylresorcinol synthases expressed in *Sorghum bicolor* root hairs play an essential role in the biosynthesis of the allelopathic benzoquinone sorgoleone. *Plant Cell*, **22**, 867-887.

Dai, X., Wang, G., Yang, D.S., Tang, Y., Broun, P., Marks, M.D., Sumner, L.W., Dixon, R.A. and Zhao, P.X. (2010). TrichOME: a comparative omics database for plant trichomes. *Plant Physiol.* **152**, 44-54.

Dayanandan, P. and Kaufman, P.B. (1976). Trichomes of *Cannabis sativa* L. (Cannabaceae). *Am. J. Bot.* 63, 578-591.

Dehesh, K., Edwards, P., Hayes, T., Cranmer, A.M. and Fillatti, J. (1996). Two novel thioesterases are key determinants of the bimodal distribution of acyl chain length of *Cuphea palustris* seed oil. *Plant Physiol.* **110**, 203-210.

Dewick, P.M. (2009). *Medicinal Natural Products: A Biosynthetic Approach* 3rd Edition. John Wiley & Sons.

Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K. and Pikaard, C.S. (2006). Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* **45**, 616-629.

Elsohly, M.A. and Slade, D. (2005). Chemical constituents of marijuana: The complex mixture of natural cannabinoids. *Life Sci.* **78**, 539-548.

Fellermeier, M. and Zenk, M.H. (1998). Prenylation of olivetolate by a hemp transferase yields cannabigerolic acid, the precursor of tetrahydrocannabinol. *FEBS Lett.* **427**, 283-285.

Fellermeier, M., Eisenreich, W., Bacher, A. and Zenk, M.H. (2001). Biosynthesis of cannabinoids. Incorporation experiments with ^{13}C-labeled glucoses. *Eur. J. Biochem.* **268**, 1596-1604.

Gaoni, Y. and Mechoulam, R. (1964). Isolation, structure, and partial synthesis of an active constituent of hashish. *J. Am. Chem. Soc.* **86**, 1646-1647.

Gershenzon, J., McCaskill, D., Rajaonarivony, J.I., Mihaliak, C., Karp, F. and Croteau, R. (1992). Isolation of secretory cells from plant glandular trichomes and their use in biosynthetic studies of monoterpenes and other gland products. *Anal. Biochem.* **200**, 130-138.

Ghosh, R., Chhabra, A., Phatale, P.A., Samrat, S.K., Sharma, J., Gosain, A., Mohanty, D., Saran, S. and Gokhale, R.S. (2008). Dissecting the functional role of polyketide synthases in *Dictyostelium discoideum*: biosynthesis of the differentiation regulating factor 4-methyl-5-pentylbenzene-1,3-diol. *J. Biol. Chem.* **283**, 11348-11354.

Gray, D. and Prestage, S. (1999). Fresh tomato specific fluctuations in the composition of lipoxygenase-generated C6 aldehydes. *Food Chem.* **64**, 149-155.

Horper, W. (1996). Biosynthesis of primin and miconidin and its derivatives. *Phytochemistry* **41**, 451-456.

Hu, Y., Gai, Y., Yin, L., Wang, X., Feng, C., Feng, L., Li, D., Jiang, X.-N. and Wang, D.-C. (2010). Crystal structures of a *Populus tomentosa* 4-coumarate:CoA ligase shed light on its enzymatic mechanisms. *Plant Cell* **22,** 3093-3104.

Izzo, A.A., Borrelli, F., Capasso, R., Di Marzo, V. and Mechoulam, R. (2009). Non-psychotropic plant cannabinoids: new therapeutic opportunities from an ancient herb. *Trends Pharmacol. Sci*. **30**, 515-527.

Kienow, L., Schneider, K., Bartsch, M., Stuible, H.-P., Weng, H., Miersch, O., Wasternack, C. and Kombrink, E. (2008). Jasmonates meet fatty acids: functional analysis of a new acyl-coenzyme A synthetase family from *Arabidopsis thaliana*. *J. Exp. Bot.* **59**, 403-19.

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

Kroumova, A.B. and Wagner, G.J. (2003). Different elongation pathways in the biosynthesis of acyl groups of trichome exudate sugar esters from various solanaceous plants. *Planta* **216**, 1013-1021.

Kroumova, A.B., Xie, Z. and Wagner, G.J. (1994). A pathway for the biosynthesis of straight and branched, odd- and even-length, medium-chain fatty acids in plants. *Proc. Natl. Acad. Sci. USA*, **91**, 11437-11441.

Lange, B.M., Wildung, M.R., Stauber, E.J., Sanchez, C., Pouchnik, D. and Croteau, R. (2000). Probing essential oil biosynthesis and secretion by functional evaluation of expressed sequence tags from mint glandular trichomes. *Proc. Natl. Acad. Sci. USA*, **97**, 2934-2939.

Lanyon, V.S., Turner, J.C. and Mahlberg, P.G. (1981). Quantitative analysis of cannabinoids in the secretory product from capitate-stalked glands of *Cannabis sativa* L. (Cannabaceae). *Bot. Gaz.* **142,** 316-319.

Lingner, T., Kataya, A.R., Antonicelli, G.E., Benichou, A., Nilssen, K., Chen, X.-Y., Siemsen, T., Morgenstern, B., Meinicke, P. and Reumann, S. (2011). Identification of novel plant peroxisomal targeting signals by a combination of machine learning methods and in vivo subcellular targeting analyses. *Plant Cell* **23**, 1556-72.

Marks, M.D., Tian, L., Wenger, J.P., Omburo, S.N., Soto-Fuentes, W., He, J., Gang, D.R., Weiblen, G.D. and Dixon, R.A. (2009). Identification of candidate genes affecting Δ^9-tetrahydrocannabinol biosynthesis in *Cannabis sativa*. *J. Exp. Bot.* **60**, 3715-3726.

Matoba, T., Hidaka, H., Narita, H., Kitamura, K., Kaizuma, N. and Kito, M. (1985). Lipoxygenase-2 isozyme is responsible for generation of *n*-hexanal in soybean homogenate. *J. Agric. Food Chem.* **33**, 852-855.

Mechoulam, R., Parker, L.A. and Gallily, R. (2002). Cannabidiol: an overview of some pharmacological aspects. *J. Clin. Pharmacol.* **42**, 11S-19S.

Mechoulam, R. (2005). Plant cannabinoids: a neglected pharmacological treasure trove. *Br. J. Pharmacol*. **146**, 913-915.

Meisel, L., Fonseca, B., Gonzaelz, S., Baeza-Yates, R., Cambiazo, V., Campos, R., Gonzalez, M., Orellana, A., Retamales, J. and Silva, H. (2005). A rapid and efficient method for purifying high quality total RNA from peaches (*Prunus persica*) for functional genomics analyses. *Biol. Res.* **38,** 83-88.

Nagel, J., Culley, L., Lu, Y., Liu, E., Matthews, P.D., Stevens, J.F. and Page, J.E. (2008). EST analysis of hop glandular trichomes identifies an *O*-methyltransferase that catalyzes the biosynthesis of xanthohumol. *Plant Cell*, **20**, 186-200.

Nelson, B.K., Cai, X. and Nebenführ, A. (2007). A multicolored set of *in vivo* organelle markers for co-localization studies in Arabidopsis and other plants. *Plant J.* **51**, 1126-1136.

Perera, M.A., Choi, S.-Y., Wurtele, E.S. and Nikolau, B.J. (2009). Quantitative analysis of short-chain acyl-coenzymeAs in plant tissues by LC-MS-MS electrospray ionization method. *J. Chromatogr.* **877**, 482-488.

Phillips, M.A., León, P., Boronat, A. and Rodríguez-Concepción, M. (2008). The plastidial MEP pathway: unified nomenclature and resources. *Trends Plant Sci.* **13**, 619-623.

Pichersky, E. and Gang, D.R. (2000). Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends Plant Sci.* **5**, 439-445.

Raharjo, T.J., Chang, W.T., Verberne, M.C., Peltenburg-Looman, A.M., Linthorst, H.J. and Verpoorte, R. (2004). Cloning and over-expression of a cDNA encoding a polyketide synthase from *Cannabis sativa*. *Plant Physiol. Biochem* **42**, 291-297.

Reumann, S. (2004). Specification of the peroxisome targeting signals type 1 and type 2 of plant peroxisomes by bioinformatics analyses. *Plant Physiol.* **135**, 783-800.

Rowan, D.D., Allen, J.M., Fielder, S. and Hunt, M.B. (1999). Biosynthesis of straight-chain ester volatiles in red delicious and granny smith apples using deuterium-labeled precursors. *J. Agric. Food. Chem.* **47**, 2553-2562.

Schmelz, S. and Naismith, J.H. (2009). Adenylate-forming enzymes. *Curr. Op. Struct. Biol.* **19**, 666-671.

Schneider, K., Kienow, L., Schmelzer, E., Colby, T., Bartsch, M., Miersch, O., Wasternack, C., Kombrink, E. and Stuible, H.-P. (2005). A new type of peroxisomal acyl-coenzyme A synthetase from *Arabidopsis thaliana* has the catalytic capacity to activate biosynthetic precursors of jasmonic acid. *J. Biol. Chem.* **280**, 13962-13972.

Shockey, J. and Browse, J. (2011). Genome-level and biochemical diversity of the acyl-activating enzyme superfamily in plants. *Plant J.* **66**, 143-160.

Shockey, J.M., Fulda, M.S. and Browse, J.A. (2002). Arabidopsis contains nine long-chain acyl-coenzyme a synthetase genes that participate in fatty acid and glycerolipid metabolism. *Plant Physiol.* **129**, 1710-1722.

Shockey, J.M., Fulda, M.S. and Browse, J. (2003). Arabidopsis contains a large superfamily of acyl-activating enzymes. Phylogenetic and biochemical analysis reveals a new class of acyl-coenzyme A synthetases. *Plant Physiol.* **132**, 1065-1076.

Sirikantaramas, S., Morimoto, S., Shoyama, Y., Ishikawa, Y., Wada, Y., Shoyama, Y. and Taura, F. (2004). The gene controlling marijuana psychoactivity: molecular cloning and heterologous expression of Δ^1-tetrahydrocannabinolic acid synthase from *Cannabis sativa* L. *J. Biol. Chem.* **279**, 39767-39774.

Souza, C. de A., Barbazuk, B., Ralph, S.G., Bohlmann, J., Hamberger, B. and Douglas, C.J. (2008). Genome-wide analysis of a land plant-specific *acyl:coenzyme A synthetase* (*ACS*) gene family in *Arabidopsis*, poplar, rice and *Physcomitrella*. *New Phytol*. **179**, 987-1003.

Sparkes, I.A., Runions, J., Kearns, A. and Hawes, C. (2006). Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nat. Protocol.* **1**, 2019-2025.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731-2739.

Taura, F., Sirikantaramas, S., Shoyama, Y., Yoshikai, K., Shoyama, Y. and Morimoto, S. (2007). Cannabidiolic-acid synthase, the chemotype-determining enzyme in the fiber-type *Cannabis sativa*. *FEBS Lett.* **581**, 2929-2934.

Taura, F., Tanaka, S., Taguchi, C., Fukamizu, T., Tanaka, H., Shoyama, Y. and Morimoto, S. (2009). Characterization of olivetol synthase, a polyketide synthase putatively involved in cannabinoid biosynthetic pathway. *FEBS Lett.* **583**, 2061-2066.

Townsend, C.A., Christensen, S.B. and Trautwein, K. (1984). Hexanoate as a starter unit in polyketide biosynthesis. *J. Am. Chem. Soc.***106**, 3868-3869.

Wang, G., Tian, L., Aziz, N., Broun, P., Dai, X., He, J., King, A., Zhao, P.X. and Dixon, R.A. (2008). Terpene biosynthesis in glandular trichomes of hop (*Humulus lupulus* L.). *Plant Physiol*. **148**, 1254-1266.

Xiao, S. and Chye, M.-L. (2011). New roles for acyl-CoA-binding proteins (ACBPs) in plant development, stress responses and lipid metabolism. *Prog. Lipid Res.* **50**, 141-151.

Xu, Z., Metsä-Ketelä, M. and Hertweck, C. (2009). Ketosynthase III as a gateway to engineering the biosynthesis of antitumoral benastatin derivatives. *J. Biotechnol.* **140**, 107-113.

Yabe, K. and Nakajima, H. (2004). Enzyme reactions and genes in aflatoxin biosynthesis. *Appl. Microbiol. Biotechnol.* **64**, 745-755.

FIGURE LEGENDS

Figure 1. The cannabinoid biosynthetic pathway leading to the major cannabinoids, Δ^9-tetrahydrocannabinolic acid (THCA) and cannabidiolic acid (CBDA). The carbon chain of the fatty acid precursor hexanoate is labeled in bold to show its incorporation into cannabinoids.

Figure 2. LC MS/MS analysis of hexanoyl-CoA in female flowers of the hemp cultivar 'Finola'.

(A) Hexanoyl-CoA detected using m/z 866.2 > 359.2 and benzoyl-CoA internal standard detected with m/z 872.2 > 365.1.

(B) Hexanoyl-CoA in cannabis flower extract detected using m/z 866.2 > 359.2.

Figure 3. LC-MS/MS analysis of CsAAE1 and CsAAE3 enzyme assays producing hexanoyl-CoA.

Purified recombinant enzymes were assayed with hexanoate, CoA, Mg^{2+} and ATP. Products were identified by MRM transitions and retention time compared to an authentic hexanoyl-CoA standard. Negative controls with boiled enzymes show the lack of hexanoyl-CoA synthesis.

Figure 4. Substrate specificities of CsAAE1 and CsAAE3.

Diverse carboxylic acid substrates were tested with purified recombinant CsAAE1 or CsAAE3, CoA, Mg^{2+} and ATP. The amount of unconsumed ATP remaining at completion

was quantified using a luciferase-based system (Schneider *et al.*, 2005). Values represent the percent difference of ATP depletion compared to control reactions that contained no substrate. C2 to C20 indicate the number of carbons in saturated fatty acids; iso denotes branched chain substrates. Error bars represent the percent error of the ratio, n=3.

Figure 5. qRT-PCR analysis of *CsAAE1, CsAAE3* and *CBDAS* expression in different tissues of the hemp cultivar 'Finola'.

Gene expression values relative to *actin* were plotted as fold differences compared to leaves, with leaf expression assigned a value of 1. Insets depict gene expression in female flowers with and without trichomes, with values also indicated as fold differences compared to leaves. R, roots; S, stems; L, leaves; FF+, female flowers with trichomes; FF-, female flowers with trichomes removed by the Beadbeater method; T, trichomes; MF, male flower. Values are mean ± SD, n=3.

Figure 6. Subcellular localization of CsAAE1 and CsAAE3.

Cannabis AAEs were transiently expressed as N-terminal YFP fusions in *Nicotania benthamiana* leaves.

(A) YFP:CsAAE1. (B) OLS:CFP co-expressed in the same cells as YFP:CsAAE1. (C) Overlay of A and B. (D) YFP:CsAAE3. (E) PX-CK peroxisome marker co-expressed in the same cells as YFP:CsAAE3. (F) Overlay of D and E.

Figure 7. Analysis of AAE clade II homologs from the plant kingdom.

(A) Maximum likelihood inference of the phylogenetic relationship of CsAAE1 and representative plant clade II homologs. Branch point bootstrap values were calculated with 1000 replicates, and values less than 60 are not shown. At, *Arabidopsis thaliana*; Cr, *Chlamydomonas reinhardtii*; Cs, *Cannabis sativa*; Gm, *Glycine max*; Mt, *Medicago truncatula*; Os, *Oryza sativa*; Pp, *Physcomitrella patens*; Pt, *Populus trichocarpa*; Rc, *Ricinus communis*; Sm, *Selaginella moellendorffii*; Vc, *Volvox carteri*; Vv, *Vitis vinifera;* Zm, *Zea mays*. (B) Alignment of the C-terminal amino acid residues of clade II AAE proteins. The e-values calculated using a PTS1 signal prediction algorithm (www.peroxisomedb.org) are listed adjacent to each sequence.

Table 1. Quantity of hexanoyl-CoA and CBDA in various hemp 'Finola' tissues.

Tissue	hexanoyl-CoA pmoles g^{-1} FW[a]	CBDA µmoles g^{-1} FW[a]
Female flowers	15.5 ± 5.7	2.4 ± 0.8
Leaves	4.0 ± 0.6	0.5 ± 0.2
Stems	2.2 ± 1.2	0.05 ± 0.04
Roots	1.5 ± 0.7	0.004 ± 0.001

[a]Values represent mean ± standard deviation (n= 4)

Table 2. Transcripts in a cannabis trichome-specific cDNA library as determined by EST sequencing.

Description	UniProt #	# ESTs	% ESTs
Cannabinoid Pathway			
Olivetol synthase / polyketide synthase	B1Q2B6	339	3.70
Tetrahydrocannabinolic acid synthase	Q8GTB6	48	0.52
CsAAE1	B9RE68	42	0.46
Fatty Acid Pathway			
Delta12-oleic acid desaturase	Q6RXX0	62	0.68
Delta12-oleic acid desaturase	Q41305	53	0.58
Acetoacetyl-CoA thiolase	Q2L8A7	28	0.31
Lipoxygenase	C4NZX3	25	0.27
Delta12-oleic acid desaturase	Q6RXX0	22	0.24
Delta12-oleic acid desaturase	Q6RXX0	22	0.24
Hydroperoxide lyase	B9IAH7	3	0.03
MEP Pathway			
4-hydroxy-3-methylbut-2-enyl diphosphate reductase	B9RZD3	185	2.02
Isopentenyl pyrophosphate isomerase	Q6EJD1	48	0.52
4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	A9ZN14	33	0.36
1-deoxy-D-xylulose-5-phosphate synthase	B9RSX4	15	0.16
1-deoxy-D-xylulose-5-phosphate reductoisomerase	B9RB24	15	0.16
2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	Q94A03	1	0.01
4-(Cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase	A9ZN11	6	0.07
2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	Q0Q5B3	3	0.03
Terpenoid Biosynthesis			
Limonene synthase	A7IZZ1	65	0.71
α-pinene synthase	A7IZZ2	38	0.41
Monoterpene synthase	B6SCF4	35	0.38
Sesquiterpene synthase	B6SCF5	31	0.34
Sesquiterpene synthase	B6SCF5	27	0.29

The function of the contigs was predicted using BLASTX comparison to the UniProt database (Nov 29, 2009 version).

Table 3. Acyl-activating enzymes identified in the cDNA library from cannabis trichome secretory cells.

Name	Contig size	Closest Arabidopsis homolog, accession	% Identity
CsAAE1	42	AtAAE17, At5g23050	65
CsAAE2[a]	9	AtLACS4, At4g23850	74
CsAAE3	5	At4CL-like, At4g05160	73
CsAAE4[a]	2	AtAAE18, At1g55320	66
CsAAE5	2	AtAAE6, At5g16340	63
CsAAE6	1	AtACN1, At3g16910	74
CsAAE7	1	AtAAE2, At2g17650	64
CsAAE8	1	AtAAE3, At3g48990	76
CsAAE9	1	AtAAE1, At1g20560	72
CsAAE10	1	AtAAE13, At3g16170	72
CsAAE11[a]	1	4CL1, At1g51680	60
CsAAE12[b]	n/a	AtAAE17, At5g23050	67
CsAAE13[b]	n/a	AtAAE18, At1g55320	68
CsAAE14[b]	n/a	AtAAE18, At1g55320	69
CsAAE15[b]	n/a	ACS, At5g36680	79

[a] cDNA was not full length. Identity scores were calculated using the incomplete amino acid sequence.

[b] Sequence identified in the *Cannabis sativa* genome (van Bakel *et al.*, 2011).

Table 4. Kinetic properties of CsAAE1 and CsAAE3.

Substrate	**CsAAE1**			**CsAAE3**		
	K_m	V_{max} (pKat)	k_{cat} (s^{-1})	K_m	V_{max} (pKat)	k_{cat} (s^{-1})
CoA	0.26 ± 0.05 µM	-	-	0.16 ± 0.01 µM	-	-
butanoate	>10 mM	-	-	>10 mM	-	-
hexanoate	3.7 ± 0.7 mM	6.8 ± 0.7	2.0	261 ± 37 µM	1.8 ± 0.05	57.6
decanoate	1.7 ± 0.2 mM	1.8 ± 0.7	0.5	16.1 ± 5.8 µM	1.6 ± 0.1	10.0
palmitate	n.d.[a]	-	-	1.3 ± 0.5 µM	0.4 ± 0.01	2.4

[a] not determined due to lack of catalytic activity

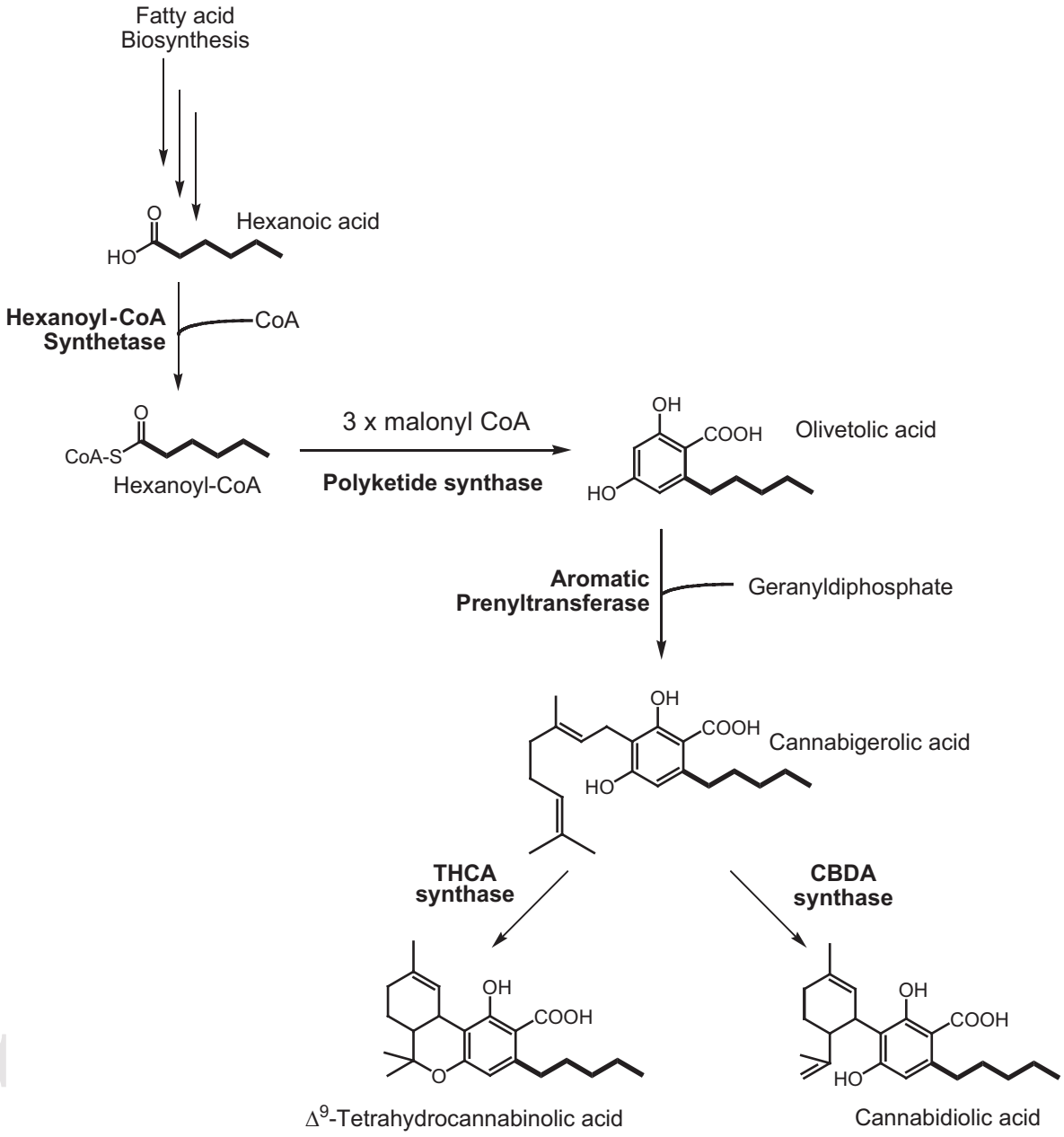
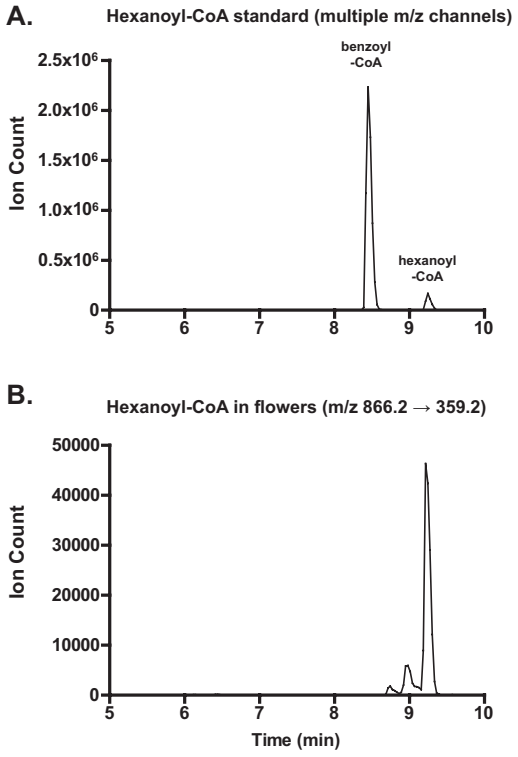
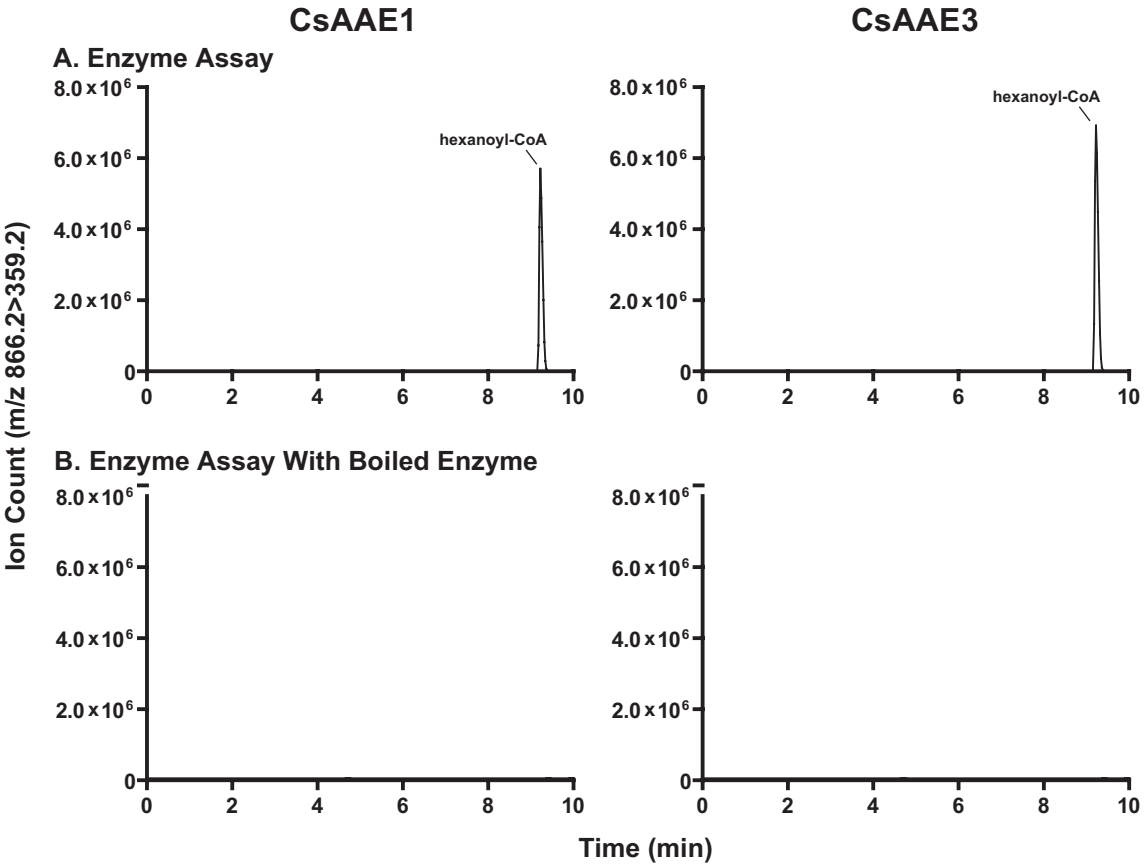


Figure 2





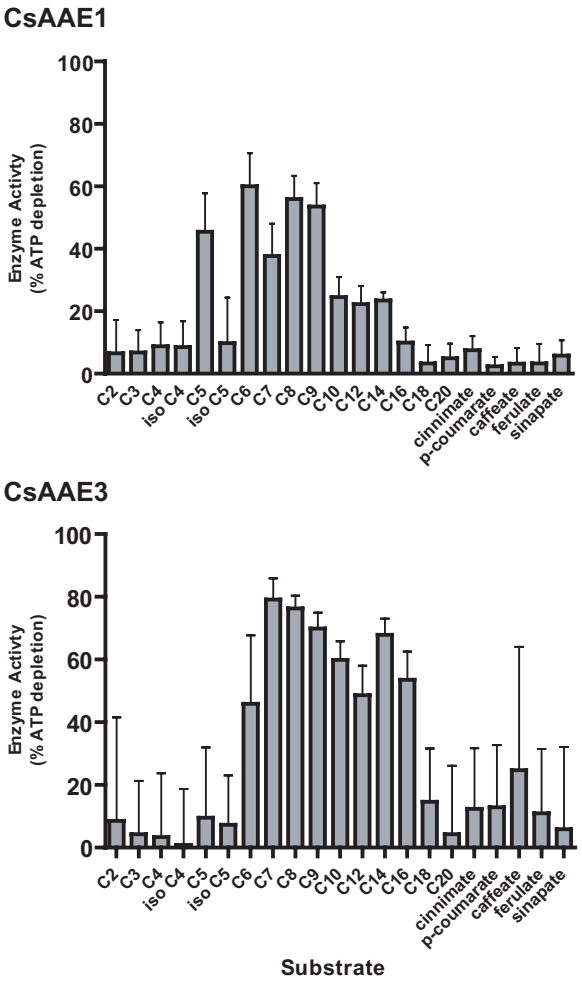


Figure 5

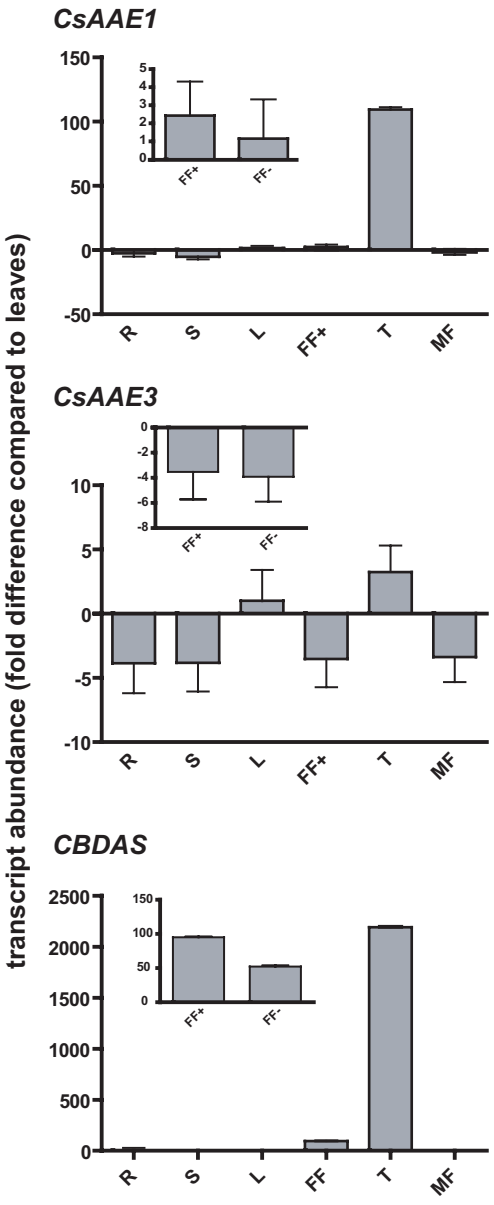
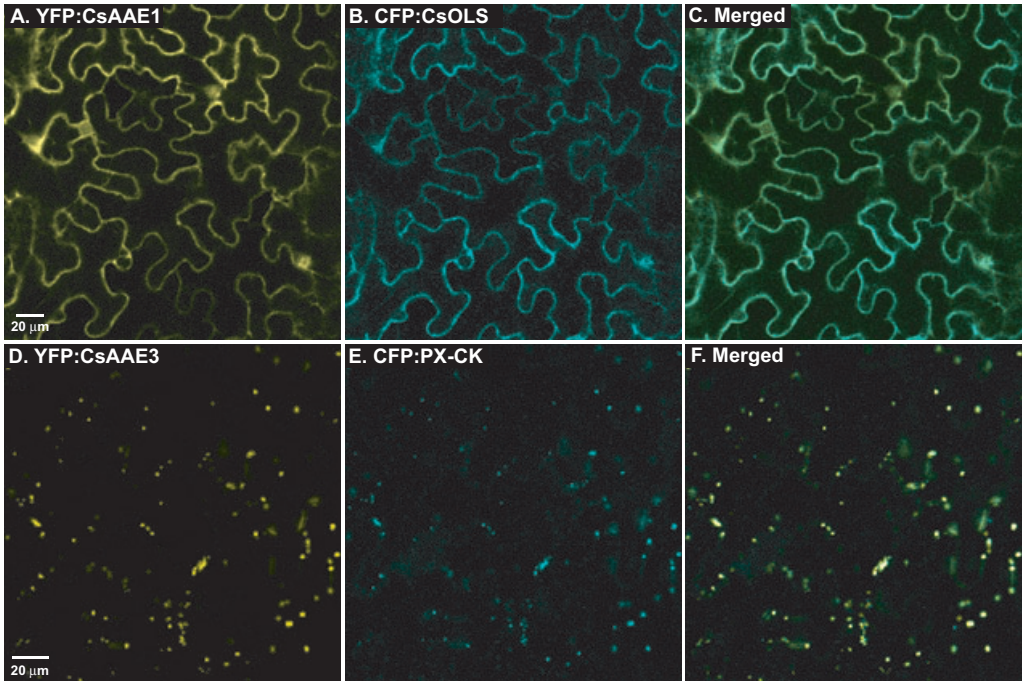
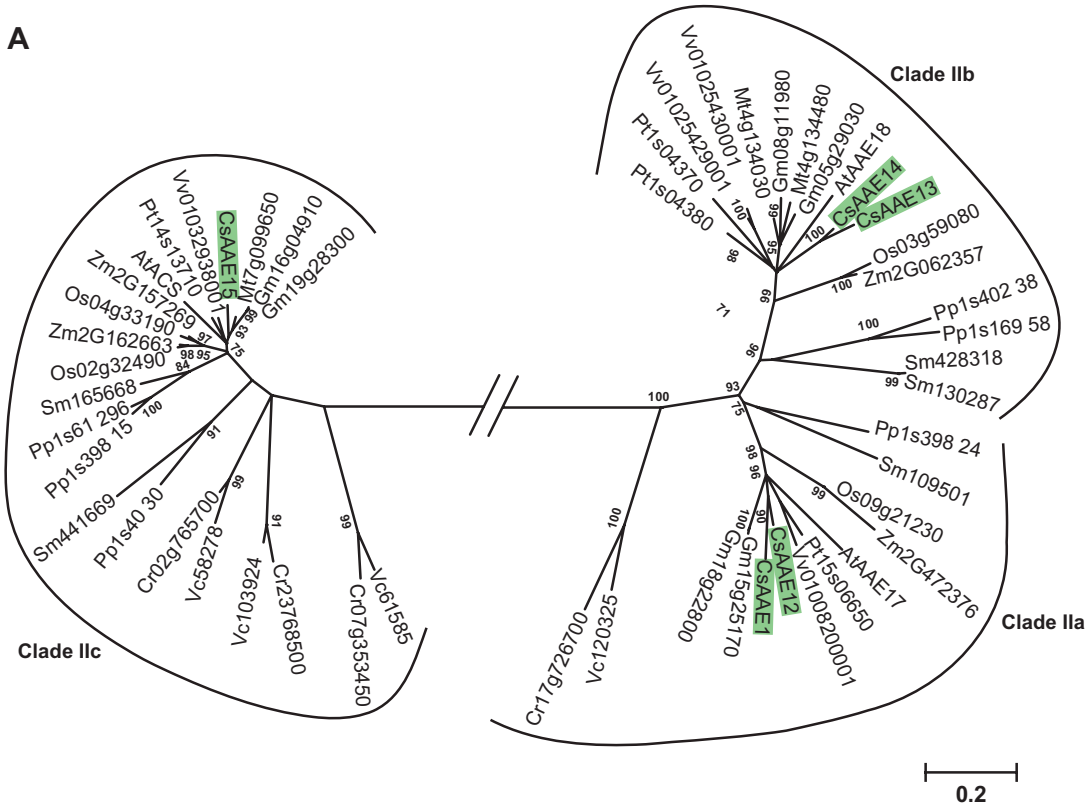


Figure 6



A



B

		e-value
CsAAE1	QFSHFE	0.35
Gm15g25170	QQLSETNQNSKI	0.25
Gm18g22800	QQLSETNQNSKI	0.25
Sm109501	LRSQSTKRTSLL	0.15
Zm2g472376	FSQAAQAKHSKI	0.12
AtAAE18	IKQELLSLRSRI	0.11
Pp1S398_24	YGPVTCPKKSSL	0.09
Vv01008200001	QQLAQLSSTSKF	0.06
CsAAE12	QQFTQLDKSSKI	0.04
CsAAE13	QLNREFSIQSKI	0.04
AtAAE17	QQLTQTGLNSKL	0.03
CsAAE14	QIKHELSAHSRI	0.03
Vv01025430001	QIKHELSVWSRI	0.03
Pp1s169_58	AKLLYSAPQSRL	0.02
Pt1s04370	QMKKELSVRSKI	0.02
Vv01025429001	QIKHELSVRSRI	0.02
Os09g21230	KEFTQQPKHSKI	0.01
Pp1s402_38	AKIVNSKQRSRL	0.01
Pt15s06650	QQLSQQDQNSKL	0.01
Pt1s04380	QMKHELSVRSKI	0.01
Cre17g726700	RDELLKQGRSRL	0.00
Gm05g29030	QMKLQLSVQSRL	0.00
Gm08g11980	QMKRELSIQSRL	0.00
Mt4g134030	QMKHELSVHSRL	0.00
Mt4g134480	QMKHELSVQSRL	0.00
Os03g59080	QLSKELSNRSKL	0.00
Sm130287	RVLRAQMARSKL	0.00
Sm428318	RVLRAQMARSKL	0.00
Vc120325	DMLLQLPGQAKL	0.00
Zm2g062357	QLAQELSNRSKL	0.00