Acyl-ACP thioesterases from *Camelina sativa*: Cloning, enzymatic characterization and implication in seed oil fatty acid composition

Manuel Fernando Rodríguez Rodríguez, Joaquín J. Salas*, Rafael Garcés and Enrique Martínez-Force

Author affiliations: Instituto de la Grasa, CSIC, Avda Padre García Tejero 4, 41012 Seville, Spain.

Corresponding author: Joaquín J. Salas
emforce@ig.csic.es
Tel: 34-954611550; Fax: 34-954616790

Abstract

Acyl-acyl carrier protein (ACP) thioesterases are intraplastidial enzymes that terminate *de novo* fatty acid biosynthesis in the plastids of higher plants by hydrolyzing the thioester bond between ACP and the fatty acid synthesized. Free fatty acids are then esterified with coenzyme A prior to being incorporated into glycerolipids. Acyl-ACP thioesterases are classified into two families, FatA and FatB, which differ in their amino acid sequence and substrate specificity. Here, the FatA and FatB thioesterases from *Camelina sativa* seeds, a crop of interest in plant biotechnology, were cloned, sequenced and characterized. The mature proteins encoded by these genes were characterized biochemically after they were heterologously expressed in *Escherichia coli* and purified. *Camelina sativa* contained three different alleles of both the FatA and FatB genes. These genes were expressed most strongly in expanding tissues in which lipids are very actively synthesized, such as developing seed endosperm. The *Cs*FatA enzyme displayed high catalytic efficiency on oleoyl-ACP and *Cs*FatB acted efficiently on palmitoyl-ACP. The contribution of these two enzymes to the synthesis of *Camelina sativa* oil was discussed in the light of these results.

Keywords: *Camelina sativa*, acyl-ACP thioesterases, FatA, FatB, fatty acid bio-synthesis.

Abbreviations:

ACP, acyl carrier protein, UFA, unsaturated fatty acids; SFA saturated fatty acids.

1. Introduction

Camelina sativa is a member of the Brasicaceae family (Al-Shehbaz, 1987) that has been cultivated in Europe since the Bronze Age (Bouby, 1998). *Camelina* accumulates high levels of oil in its seeds, representing between 15.5% and 41.7% of the seed weight, which are rich in ω-3 and ω-6 fatty acids. Typically, the major fatty acid species in *Camelina* oil are palmitic acid (16:0, 6.8% of the total fatty acids), stearic acid (18:0, 2.7%), oleic acid (18:1, 16.7%), linoleic acid (18:2, 21.9%), linolenic acid (18:3, 29.3%), gondoic acid (20:1, 13.9%) and erucic acid (22:1, 2.8%: Rodríguez-Rodríguez et al., 2013). The high percentage of polyunsaturated fatty acids in *Camelina* oil makes it unstable and inappropriate as a source of biodiesel or as a biolubricant stock (Fröhlich and Rice, 2005; Ciubota-Rosie et al., 2013).

In higher plants, *de novo* fatty acid biosynthesis takes place in the plastids and it is catalyzed by the fatty acid synthase complex (FAS), which produces aliphatic fatty acids through the sequential elongation of acyl-acyl carrier protein (ACP) derivatives (Pidkowitc et al., 2007). The main product of the FAS complexes are 16:0-ACP and 18:0-ACP. The activity of the intraplastidial stearoyl-ACP desaturase enzyme (SAD) desaturates the latter intermediate to produce oleoyl-ACP (18:1-ACP: Shanklin and Cahoon, 1998). These acyl-ACPs are exported out of the plastids to be incorporated into the extraplastidial pathways of glycerolipid synthesis in the endoplasmic reticulum (Ohlrogge et al., 2000). Fatty acid export involves the hydrolysis of the thioester bond between ACP and the acyl group, a reaction catalyzed by acyl-ACP thioesterases (Voelker et al., 1997). The products of this reaction are ACP, that is recycled for the further synthesis of acyl-ACPs, and free fatty acids that are quickly condensed with coenzyme A by long-acyl-CoA-synthetase (LACS: Koo et al., 2004; Aznar-Moreno et al., 2014).

According to their sequence and acyl-ACP preference, acyl-ACP thioesterases are usually categorized into two groups, FatA and FatB (Jones et al., 1995). FatA-type proteins show high substrate specificity towards 18:1-ACP, whereas FatB enzymes are classified into further two subclasses: FatB1, with preference towards long-chain saturated acyl-ACPs, especially 16:0-ACP (Dörmann et al., 1995); and FatB2 that prefer short/medium-chain saturated acyl-ACPs. FatA and FatB1 enzymes are present in all plants, whereas FatB2 enzymes are only found in species that accumulate C8-C14 fatty acids in their seed oils (Voelker et al., 1997).

The specificity of the acyl-ACP thioesterases is important in defining the fatty acid profile of seed oil and thus, these enzymes are relevant targets to manipulate the fatty acid composition of seed oils. For example, seeds of a double FatA *Arabidopsis thaliana* mutant in which the expression of both genes was reduced had less triacylglycerides and a modified fatty acid profile (Moreno-Pérez et al., 2012). Furthermore, a FatB *Arabidopsis thaliana* knock-out mutant had approximately 45% lower palmitic acid content, a similar reduction to that produced by FatB silencing with a miRNA (Bonaventure et al., 2003; Belide et al., 2012).

There is currently increased interest in *Camelina sativa* as a model plant for applied crop research because it is an oil crop that can be easily transformed using floral dip infiltration under vacuum (Lu and Kang, 2008), and it is genetically similar to *Arabidopsis thaliana* (Collins-Silva et al, 2011). Moreover, this plant has several agronomic advantages, such as a short vegetative period of about 90-120 days (Putnam et al., 1993), low nutritional requirements (Bramm et al., 1990; Zubr, 1997) and the secretion of phytoalexins that inhibit the development of pathogens (Lovett and Jackson, 1980). In the present work, we have cloned and characterized two acyl-ACP thioesterases from *Camelina sativa*, *Cs*FatA and *Cs*FatB. Furthermore, we have studied their distribution in the plant and their phylogenetic relationship with thioesterases from other species. Finally, the contribution of these enzymes to oil synthesis in *Camelina*, as well as their possible biotechnological applications, are discussed in the light of our results.

2. Results and Discussion

2.1 Isolation and sequence analysis of acyl-ACP thioesterases from Camelina sativa

Using two different degenerate primer pairs, Deg_FatA_F plus Deg_FatA_R and Deg_FatB_F plus Deg_FatB_R (Table 1), DNA fragments of 458 and 633 bp were amplified from developing *Camelina sativa* seed cDNA, corresponding to internal fragments of the *CsFatA* and *CsFatB* genes, respectively. Afterwards, we obtained the full-length *CsFatA* cDNA clone of 1110 bp and *CsFatB* cDNA clone of 1251 bp by RACE, as described in the Material and Methods (Table 1). The alignment of the deduced amino acid sequences showed a high degree of identity with the internal coding

regions of known FatA and FatB thioesterase protein sequences. Moreover, specific primers were designed from these sequences to amplify all possible *CsFatA* alleles (CsFatASphI_F3 and CsFatAXmaI_R3, Table 1), identifying three different alleles: *CsFatA1*, 1110 bp (GenBank accession number AFQ60947.1); *CsFatA2*, 1110 bp (GenBank accession number AFQ60948.1); and *CsFatA3*, 1107 bp (GenBank accession number AFQ60946.1). Similarly, another pair of primers were designed based on the *CsFatB* sequences (CsFatB_F3 and CsFatB_R3, Table 1), identifying the same number of alleles for this gene also: *CsFatB1*, 1251 bp (GenBank accession number AFQ60949.1); *CsFatB2*, 1251 bp (GenBank accession number AFQ60951.1); and *CsFatB3*, 1251 bp (GenBank accession number AFQ60950.1).

The *CsFatA1*, *CsFatA2* and *CsFatA3* open reading frames (ORFs) identified were predicted to encode proteins of 369, 369 and 368 amino acid residues (Fig. 1), corresponding to calculated molecular masses of 41.47, 41.49 and 41.40 kDa, and with pI values of 7.44, 7.41 and 7.19, respectively. Despite the fact that *CsFatA1*, *CsFatA2* and *CsFatA3* sequences do not appear to contain the consensus sequence for translation initiation in plants (Joshi et al., 1997), they conserve the T at the position +4 from the ATG, which is required for fidelity of translation initiation (Kozak, 1991). The ORFs of *CsFatB1*, *CsFatB2* and *CsFatB3* identified were each predicted to encode proteins of 416 amino acid residues (Fig. 2), with estimated molecular masses of 46.15, 46.11 and 46.11, and similar pI values of 8.83, 8.82 and 8.82 respectively. The nucleotides around the methionine at the start codon of *CsFatB1*, *CsFatB2* and *CsFatB3* resemble the consensus sequence for translation initiation in plants (Joshi et al., 1997). The papain-like catalytic triad required for the reaction catalyzed by thioesterase enzymes is formed by asparagine, histidine and glutamine (Mayer and Shanklin, 2007). These residues were conserved in all the alleles identified, occupying the position Asn-273, His275 and Gln311 in *Cs*FatA and Asn-318, His-320 and Gln-384 in *Cs*FatB (Fig. 1 and Fig. 2).

WoLF PSORT was used to identify the signal peptide and the cleavage site where the mature protein begins, a program that predicts protein subcellular localization (Horton et al., 2007), with the information available for acyl-ACP thioesterases from *Arabidopsis thaliana* (Dörmann et al., 1995), *Helianthus annuus* (Martínez-Force et., al, 2000; Serrano-Vega et al., 2005) *Gossypium hirsutum* (Huynh et al., 2002), *Diploknema butyracea* (Jha et al., 2006), *Madhuca longifolia* (Ghosh and Sen, 2007), *Ricinus communis* (Sánchez-García et al., 2010) and *Macadamia tetraphylla* (Moreno-Pérez et al., 2011). Val-55 was the best candidate to represent the N-terminal amino acid of the

mature CsFatA protein, corresponding to a signal peptide of 58 amino acid residues (Fig. 1). In the case of the *Cs*FatB protein, we considered Leu-93 to be the first amino acid of the mature enzyme, which would correspond to a signal peptide of 92 amino acids (Fig. 2). Taking into account the presence of a signal peptide in both classes of thioesterases, the mature *Cs*FatA and *Cs*FatB proteins were predicted to contain 310 and 324 amino acids, with molecular masses of 34.8 and 33.4 kDa, and pI values of 5.5 and 6.9, respectively. The *Cs*FatB enzyme has a hydrophobic domain between residues Leu-93 to His-128, which is absent in *Cs*FatA thioesterases and that is probably involved in anchorage to the membrane rather than affecting its activity or affinity for different substrates (Jones et al., 1995; Facciotti and Yuan, 1998).

A phylogenetic tree was generated for the novel *Camelina* thioesterase genes based on their deduced amino acid sequences and in relation to all other known plant thioesterase sequences (Fig. 3). Each set of three *Camelina* thioesterase proteins were situated in the FatA or FatB thioesterase group. Thioesterases from *Camelina* grouped very closely with those from *Arabidopsis thaliana*, *A. lyrata*, *Capsella rubella*, *Brassica juncea*, *B. napus* and *B. rapa*, all species that belong to the Brassicaceae family. Other closely related sequences were those from members of the Fabacea family, *Glycine max*, *Arachis hypogaea*, *Cicer arietinum* and *Medicago truncatula*.

2.2 Genomic organization of Camelina sativa *acyl-ACP thioesterase genes*

To analyze the genomic organization of the *CsFatA* and *CsFatB* genes, two genomic DNA fragments at the locus were amplified using two different primer pairs. Clones of 1214 and 1427 nucleotides were obtained and sequenced for *CsFatA* and *CsFatB*, respectively. The intron and exon organization of the three *CsFatA* and *CsFatB* alleles were found by comparing their cDNA and genomic DNA sequences (Table 2). The *CsFatA1* allele was 1743 bp long, the *CsFatA2* allele was 1745 bp and the *CsFatA3* allele was 1773 bp. All *CsFatA* alleles had six introns and thus, they each contained seven exons, which were of similar length (Table 2). Intron 3 of the *CsFatA3* allele differed most in length as it contained an insertion of 32 nucleotides at the beginning of the sequence (169 bp in i3CsFatA1; 171 bp in i3CsFatA2; and 200 bp in i3CsFatA3: Table 2). The *CsFatB1* allele was 1790 bp long, the *CsFatB2* allele 1872 bp and the *CsFatB3* allele 1870 bp. All have four introns and five exons with a high degree of identity and homology between them, except for introns 2 and 3 that display significant

differences in both sequence and length (221 bp for i2CsFatB1; 305 bp for i2CsFatB2; 314 bp for i2CsFatB3; 139 bp for i3CsFatB1; 141 bp for i3CsFatB2; and 131 bp for i3CsFatB3: Table 2).

The genetic map of *Camelina sativa*, comprised of 157 amplified fragment length polymorphisms (AFLPs) and 3 single sequence repeat markers was published in 2006 (Gehringer et al., 2006). This study showed that *C. sativa* has 20 chromosomes, a figure found only among know alloploids or plant species. A triplication of the *C. sativa* genome might have resulted from two allopolyploidy events, first resulting in tetraploidy (4n) and then hexaploidy (6n), or it could also be derived from the combination of an autotetraploid (4n) and diploid (2n) species in an autopolyploidized (6n) genome (Hutcheon et al., 2010). The alignment of our intron sequences with the recently available sequences from the *C. sativa* Genome Project (http://www.camelinadb.ca/), which is included in The Prairie Gold Project, allowed the different alleles of the thioesterase genes to be located in the *Camelina* genome: *CsFatA1*, *CsFatA2* and *CsFatA3* on chromosome 15, 19 and 1, respectively; and *CsFatB1*, *CsFatB2* and *CsFatB3* on chromosome 14, 17 and 3, respectively. Analyzing the intron sequences suggests the existence of two groups of sequences for each thioesterase. Thus, s*CsFatA1* and *CsFatA2* presented strong identity with obvious differences from *CsFatA3*. Similarly the *CsFatB3* sequences showed much more variability with respect to the other two alleles, corresponding to the complementation group of chromosomes previously reported (Gehringer et al., 2006). These results are more consistent with the *Camelina* genome being autopolyploid due to the combination of an autotetraploid and a diploid species. The diploid parent could have contributed to the *C. sativa* genome with two possible combinations, 7+7+6 or 6+6+8 giving the total of 20 chromosomes.

2.3 Fatty acid analysis of E. coli expressing Camelina sativa acyl-ACP thioesterases

Mature *CsFatA* and *CsFatB* proteins were overexpressed in *E. coli* after removing the hydrophobic domain of CsFatB to enhance its expression in the soluble fraction without affecting its activity or substrate specificity (Jones et al., 1995; Facciotti and Yuan, 1998). Only one allele of each acyl-ACP thioesterase type, *CsFatA* and *CsFatB*, was cloned into the pQE-80L vector because the *Cs*FatA alleles only have three differences in their amino acid sequence, one conservative (Asp-167-Arg) and two

semi-conservative changes (Ser-128-Gly; Asp-294-His), and the different CsFatB alleles have four purely conservative changes (Leu-134-Ile; Val-305-Phe; Lys-343-Arg; Ser-370-Ala). The amino acid residues involved in substrate recognition and those related to the hydrolase activity of CsFatA and CsFatB were identical in the different alleles (Fig. 1 and Fig. 2).

In plants, the substrate specificity of thioesterases determines the oil composition because these enzymes are involved in the export of acyl-ACP from the plastid to cytosol. In *E coli*, thioesterases cleave acyl-ACPs producing the free fatty acids necessary for regulatory signals, export or degradation (Lennen and Pfleger, 2012). The fatty acid composition of *E. coli* expressing *Camelina* acyl-ACP thioesterase genes were analyzed and compared with control cells transformed with the empty pQE-80L vector (Table 3). The expression of *Cs*FatA and *Cs*FatB produced a 45% and 68% decrease in the total fatty acid content of *E. coli*, respectively. These results showed that *C. sativa* thioesterases alter *E. coli* fatty acid metabolism, diverting the acyl chains away from the fatty acid and lipid biosynthetic pathways. These free fatty acids would later be secreted or degraded in the β-oxidation pathway (Lennen and Pfleger, 2012). The main change in the *E. coli* fatty acid composition when *Cs*FatA was expressed was the reduction in unsaturated fatty acids, mainly oleic acid (18:1ω9). This reduction was compensated for by an increase of palmitoleic acid (16:1ω7). However, the expression of *Cs*FatB caused the opposite effect, a decrease in saturated fatty acids and in particular, that of palmitic acid (16:0) that is compensated for with an increase in stearic acid (18.0).

2.4 Substrate specificity and kinetic parameters of *Camelina acyl-ACP thioesterases*

The kinetic parameters of recombinant *Cs*FatA and *Cs*FatB were investigated after purification by metal ion affinity chromatography (IMAC). This method allowed to obtain highly purified enzymes in a single step (see Fig. 5). The substrate specificity of the *Cs*FatA and *Cs*FatB enzymes was determined by assaying their activity on different acyl-ACP substrates at a constant concentration (Fig. 6). The *Cs*FatA enzyme displayed a high level of activity on unsaturated fatty acid derivatives, mainly with 18:1-ACP, and it was 14-fold less active towards 18:0-ACP. The activity of *Cs*FatB did not show important differences among the substrates assayed, although *Cs*FatB presented a preference for 16:0-ACP and 18:0-ACP. These results are similar to those

reported previously for thioesterases from other plants, such *Garcinia.mangostana* (Hawkins and Kridl, 1998), *Carthamus tinctorius* (Knutzon et al., 1992), *Brassica campestri* (Pathak et al., 2004), *Arabidopsis thaliana* (Salas and Ohlrogge, 2002), *Helianthus annuus* (Serrano-Vega, et al., 2005) and *Ricinus communis* (Sánchez-García et al., 2010).

Kinetic parameters were also calculated for both enzymes acting on different substrates, displaying similar K_m values for all of them, all in the micromolar order. These values were slightly higher than those reported previously for acyl-ACP thioesterases from *Helianthus annuus* (Serrano-Vega, et al., 2005), *Ricinus communis* (Sánchez-García et al., 2010) or *Macadamia tetraphylla* (Moreno-Perez, et al., 2011). The V_{max} of *Cs*FatA for 18:1-ACP was 92.56 nkat/mg prot, one order of magnitude higher the V_{max} found for the other substrates assayed. The K_{cat} and catalytic efficiency (K_{cat}/K_m) values were also highest for this substrate (Table 4), which is in good agreement with the kinetic parameters described for most FatAs investigated to date. Similarly, *Cs*FatB displays a typical profile of FatB enzymes, showing greater catalytic efficiency towards 16:0-ACP. These V_{max} and K_{cat} values were lower than those described by Sánchez-García et al. (2010) for *Ricinus communis*, yet they were higher than those reported for *Macadamia tetraphylla* (Moreno-Perez et al., 2011).

2.5 Expression profiles of Camelina sativa acyl-ACP thioesterases

The expression of the acyl-ACP thioesterase genes in developing seeds and vegetative tissues of *Camelina sativa* was studied by QRT-PCR. The profile of transcript accumulation was temporally regulated during the development of the embryo (Fig. 6), with the strongest expression of the *CsFatA* and *CsFatB* genes occurring in developing seeds at 12, 18 and 24 DAF, the phase of oil accumulation (He et al., 2004). In seed tissue, of *CsFatA* was always expressed more strongly than *CsFatB*, which fits well with the composition of Camelina oil in which linoleic and linolenic acids predominate, a fatty acid derived from oleic acid that is mainly exported via FatA.

The expression of these genes is significantly weaker in vegetative tissue, with the exception of *CsFatB* in leaves that could be involved in the production of saturated fatty acids used for surface lipid biosynthesis. Indeed, stronger expression of *CsFatB* in leaves was also reported in *Ricinus communis* (Sánchez-García, et al., 2010).

Nevertheless, the expression patterns observed in this analysis suggests that *Camelina sativa* acyl-ACP thioesterases are important for oil deposition in the seed.

3. Conclusions

The cloning and sequencing of *Cs*FatA and *Cs*FatB thioesterases from developing *Camelina* seeds shows that they are encoded by a single copy gene, three different alleles existing of each gene. In both cases, the differences found in the coding region between these alleles are not important, accumulating mostly single nucleotide polymorphisms (SNP), insertions and deletions in the introns. Indeed, the highly conserved papain-like catalytic triad, asparagine, histidine and glutamine, are maintained in *Cs*FatA and *Cs*FatB. The heterologous expression of these enzymes in *E. coli* produced a contrasting effect on bacterial fatty acid composition, *Cs*FatA causing a decrease in the unsaturated fatty acids, mainly 18:1ω11, and *Cs*FatB augmenting these fatty acids. The substrate specificity of these enzymes is similar to that reported previously in other plants, *Cs*FatA showing a strong preference for 18:1-ACP and *Cs*FatB for 16:0-ACP. The kinetic parameters of both enzymes differ only slightly from those described in *Helianthus annus*, *Ricinus communis* or *Macadamia tetraphylla.*

4. Materials and Methods

4.1 Biological material and growth conditions

Camelina sativa plants of the CAS-CS0 cultivar were grown in chambers at 25/15 °C (day/night) with a 16 h photoperiod and a light intensity of 250 μE $m^{-2}s^{-1}$. The roots, cotyledons and hypocotylous leaves were collected 10 days after germination, and developing *Camelina* seeds were recovered during the oil active synthesis period (6, 12, 18, 24 and 30 days after flowering -DAF). All tissues were frozen in liquid nitrogen and stored at -80 ºC until use. The *Escherichia coli* strain XL1-Blue (Stratagen) was used as the plasmid hosts for the cloning and expression of *CsFatA* and *CsFatB* genes. Bacteria were grown in LB media (1 % Bacto Tryptone, 0.5 % Yeast Extract, 1% NaCl [pH 7]) and the liquid culture was shaken vigorously at 37 °C. Plasmids were selection for with ampicillin (100 μg/μl).

4.2 mRNA preparation and cDNA synthesis

Approximately 0.25 g of each tissue from *Camelina* was ground in liquid nitrogen using a precooled sterile mortar and pestle. Total RNA was isolated using the Spectrum Plant Total RNA Kit (Sigma) and the mRNA was recovered using a GenElute mRNA Miniprep Kit (Sigma). The mRNA pellet was resuspended in 33 µl RNAase-free TE buffer (10 mM Tris-HCl, 1 mM EDTA [pH 8]) and cDNAs were synthesized using the Ready-To-Go T-Primed First-Strand Kit (Amersham Bioscience).

4.3 Cloning of genes coding for Camelina sativa acyl-ACP thioesterases

The FatA and FatB protein sequences in the public databases were aligned to identify regions of homology using the ClustalX v1.8 program (Thompson et al., 1997) and this alignment was used for primer design. Two different PCR fragments were amplified with degenerate primers corresponding to highly conserved regions of FatA (Deg_FatA_F and Deg_FatA_R) and FatB (Deg_FatB_F and Deg_FatB_R; see Table 1), using cDNA from developing *Camelina sativa* kernels as the template. The 5'- and 3'-ends of the corresponding cDNAs were obtained using the SMART-RACE cDNA Amplification Kit (Clontech) and the specific primers CsFatA_F1, CsFatA_F2, CsFatA_R1 and CsFatA_R2 for the *CsFatA* gene, and CsFatB_F1, CsFatB_F2, CsFatB_R1 and CsFatB_R2 for the *CsFatB* gene from *C. sativa* (Table 1). The PCR fragments that encode the whole proteins were cloned into the pMBL-T vector (Dominion MBL), sequenced and their identities confirmed using the BLAST software (Altschul et al., 1990). The identification of different acyl-ACP thioesterase alleles in *C. sativa* was carried out by designing two pair of primers: CsFatASphI_F3 and CsFatAXmaI_R3 for the *CsFatA* gene; and CsFatB_F3 and CsFatB_R3 for the *CsFatB* gene. These primers were used to amplify these genes from cDNA obtained from developing *Camelina sativa* CAS-CS0 seeds as the template. The different amplicons were cloned into the pSpark vector (Canvax), sequenced and their identities were confirmed using the BLAST software (Altschul et al., 1990).

Primers with internal *SphI* and *XmaI* restriction sites for *CsFatA* (CsFatASphI_F4 and CsFatAXmaI_R3) and *CsFatB* (CsFatBSphI_F4 and CsFatBXmaI_R4) were designed to amplify the coding regions of the mature proteins

by PCR without a signal peptide. The PCR products were subcloned into the *SphI-XmaI* sites of the pQE-80L expression vector (Quiagen) to produce the corresponding fusion proteins with a 6xHis tag at the N-terminus. Ligation into the correct reading frame was confirmed by sequencing. The resulting constructs were named pQE80L_CsFatA and pQE80L_CsFatB, and they were used for protein overexpression in *E. coli*.

4.4 Cloning of acyl-ACP thioesterase introns of Camelina sativa

Genomic DNA from *Camelina sativa* was isolated using the GenElute™ Plant Genomic DNA Miniprep Kit (Sigma-Aldrich, Missouri). Based on the genomic structure of the *AtFatA1* (gene model in TAIR: AT3G25110.1), *AtFatA2* (gene model in TAIR: AT4G13050.1) and *AtFatB* (gene model in TAIR: AT1G08510.1) genes from *Arabidopsis thaliana*, one pair of primers were designed for each gene to amplify the intronic regions: CsFatA_INTR_F1 and CsFatA_INTR_R1 for the *CsFatA* gene; and CsFatB_INTR_F1 and CsFatB_INTR_R1 for the *CsFatB* gene (Table 1). DNA fragments of around 1000-1300 bp were amplified from the *C. sativa* genomic DNA and cloned into the pSpark vector (Canvax) for sequencing.

4.5 Protein expression and purification

Cultures of *E. coli* XL1-Blue cells (1L) harboring pQE80L_CsFatA or pQE80L_CsFatB were grown as described above. Expression of the *Cs*FatA or *Cs*FatB enzymes from *Camelina sativa* was induced at OD_{600} 0.4 - 0.6 by adding 1mM IPTG and growing the bacteria for a further 4 h. The cells were harvested by centrifugation (10 min at 2500 g), washed with distilled water and resuspended in 10 ml of Tris-HCl 50 mM [pH 8]. The cells were pre-lysed for 30 min at room temperature in 0.5% Triton X-100, 5mM DTT, 1mM $MgCl_2$, 1mM PMSF, 20 μg/ml DNase I (Roche) and 0.2 mg/ml lysozyme (Roche) with mild agitation. The cells were then disrupted by sonication using the Digital SLPe Ultrasonic System (Branson, model 4C15) equipped with a micro tip. The cells were intermittently sonicated on ice for 30 s at 70 % amplitude with 30 s cooling intervals. The resulting disrupted cell suspension was centrifuged at 25,000 × g for 30 min. The supernatant was filtered through 0.22 μm filters and loaded onto a 2 mL Ni-NTA Agarose column (Qiagen) that was then gently agitated for 1 hour at room temperature. The resin was washed with 6 ml Wash Buffer

(50mM sodium phosphate, 300mM Sodium chloride and 20mM imidazole), and the histidine-tagged proteins were eluted from the column using 6 ml of Elution Buffer (50mM sodium phosphate, 300mM Sodium chloride and 250mM imidazole) and concentrated in 1.5 ml using 30KDa Amicon Ultra Centrifugal Filters (Millipore). The purified proteins were used to assay thioesterase activity.

4.6 Fatty acid analysis

Cultures of *E. coli* XL1-Blue cells (25 ml) harboring the empty pQE-80L or pQE80L_CsFatA or pQE80L_CsFatB were grown and the expression of the corresponding thioesterase was induced as described above. The cells were harvested by centrifugation (10 min at 2500 × g), washed in distilled water and the total fatty acid content was determined using the one-step method reported by Garcés and Mancha (1993). Volumes of 3.3 ml of methanol/toluene/dimethoxypropane/sulfuric acid (39:20:5:2) and 1.7 ml heptane were added to the pellet of bacterial cells, along with 100 – 150 µg of heptadecanoic acid as an internal standard. The mixture was heated to 80 °C for 1 h and after cooling, the upper phase containing the fatty acid methyl esters was transferred to a fresh tube, washed with 6.7% sodium sulfate and evaporated to dryness with nitrogen. The methyl esters were dissolved in an appropriate volume of heptane, and analyzed by GC using a Hewlett–Packard 6890 gas chromatography set-up (Palo Alto, CA, USA) and a Supelco SP-2380 fused-silica capillary column (30 m length, 0.25 mm i.d., 0.20 µm film thickness; Supelco, Bellefonte, PA, USA). Hydrogen was used as the carrier gas at 28 cm/s, the temperature of flame ionization detector and injector was 200 °C, the oven temperature was 170 °C and the split ratio was 1:50. The different methyl esters were identified by comparison of retention times with those from known standards.

4.7 Gel electrophoresis of proteins

Protein samples were diluted in 2x SDS-PAGE sample buffer (62.5mM Tris-HCL [pH 6.8], 2% SDS, 25% glycerol, 0.01% Bromophenol Blue, 5% 2-mercaptoethanol) and heated to 98 °C for 5 min. The samples were resolved by electrophoresis on 4-15% Mini-Protean TGX Gels (BIO-RAD) in SDS-PAGE Running Buffer (25mM Tris-Base, 192mM Glycine, 0.1% SDS), and the gels were fixed and

stained with 0.1% Coomassie R-250 in 40% ethanol and 10% acetic acid. The molecular weight standards used were Precision Plus Protein Standards (BIO-RAD).

4.8 Holo-ACP activation and preparation of the acyl-ACP substrates

Labeled acyl-ACP substrates were prepared using recombinant *E. coli* acyl-ACP synthetase and the *E. coli* ACP Histidine-tagged kindly provided by Dr John Shanklin (Brookhaven National Laboratory, Upton, NY, USA) and Dr Penny Von Wettstein-Knowles (Department of Biology, University of Copenhagen, Copenhagen, Denmark) respectively. The production of holo-ACP was carried out through the method reported by Haas et al. (2000). In summary, a volume of 0.5 mL of 207 µM apo-ACP in 50 mM Tris-HCl [pH 8.8], 1mM Coenzyme A, 5mM DTT, 10mM $MgCl_2$ was preincubated at 37 °C before adding 200 nM of *E. coli* holo-ACP synthase. The reaction was then run to completion for 4 hours at 37 °C (Lambalot and Walsh, 1997). For ACP acylation, 50 µg of recombinant *E. coli* holo-ACP was reacted with 660 MBq (approximant 0.1µmol) of [1-^{14}C] fatty acid ammonium salt ([^{3}H] fatty acid in the case of palmitoleic acid) in 100 mM Tris [pH 8.0], 5 mM ATP, 2 mM DTT, 4 mM $LiCl_2$, 10 mM $MgCl_2$ and 10 µg acyl-ACP synthetase, to a final volume of 0.5 ml. Reactions were carried out at room temperature for 3–4 h and the acyl-ACPs were purified and concentrated by ion exchange chromatography on DEAE-sepharose as described by Rock and Garwin (1979).

4.9 Acyl-ACP thioesterase assays

Thioesterase activity was assayed in 0.1 ml reactions containing 50 mM Tris–HCl [pH 8.0], 5mM DTT, 1.04–0.02 ng or 2.08–0.10 ng of recombinant CsFatA or CsFatB, respectively, and the labeled acyl-ACP substrate ranging from 0.02 to 0.08 nmol (65–2100 Bq, approximately). The reactions were carried out at room temperature for 5 min and stopped by the addition of 0.25 ml of 1 M acetic acid in 2-propanol. Unesterified fatty acids were then extracted twice with 0.3 ml hexane, and the radioactivity in the pooled organic phase was determined in a calibrated liquid scintillation counter (Rackbeta II; LKB). The data from the thioesterase assays were fitted to the Lineweaver-Burk equation by non-linear least-squares regression analysis

using OriginPro 8 software and correlated at $P < 0.005$, as determined by a Student's t-test. Both the V_{max} and K_m were derived from these curves.

4.10 Quantitative real time PCR

The cDNAs were subjected to quantitative real time PCR (QRT-PCR) using specific pairs of primers (CsFatA_qPCR_F and CsFatA_qPCR_R for the *CsFatA* gene, and CsFatB_qPCR_F and CsFatB_qPCR_R for the *CsFatB* gene: Table 1) and with the SYBR Green I (QuantiTect_ SYBR_ Green PCR Kit, Quiagen, Crawley, UK) using a MiniOpticon system (Bio-Rad). The reaction mixture was heated to 50 °C for 2 min and then to 95 °C for 15 min, before subjecting it to 40 PCR cycles of 94 °C for 15 s, 55 °C for 30 s, and 72 °C for 15 s, monitoring the resulting fluorescence. Calibration curves were constructed using sequential dilutions of cDNA from kernels obtained at 15 DAF and these curves were used to estimate the transcript content of the calibrator gene. The Livak method (Livak and Schmittgen, 2001) was applied to calculate the comparative expression between samples. The *Camelina sativa* actin gene (GenBank accession number: KJ670375) was used as an internal reference to normalize the relative amount of cDNA for each sample, amplified with the specific primers: CsActin_qPCR_F and CsActin_qPCR_R (Table 1).

5. ACKNOWLEDGEMENTS

We are grateful to Bárbara López Cordero, Mª Soledad Parra Camacho and Aranzazu González Callejas for their technical assistance. This work was supported by grants from the MINECO and FEDER (ProjectsAGL2011-23187 and BIOLUBS CC09-068).

REFERENCES

Al-Shehbaz, I.A., 1987. The genera of Alysseae (Cruciferae; Brassicaceae) in the Southeastern United States. J. Arnold Arbor. 68, 185-240.

Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403-410

Aznar-Moreno, J.A., Venegas Calerón, M., Martínez-Force, E., Garcés, R., Mullen, R., Gidda, S.K., Salas, J.J., 2014. Sunflower (Helianthus annuus) long-chain acyl-coenzyme A synthetases expressed at high levels in developing seeds. Physiologia Plantarum. 150, 363-373.

Belide, S., Robertson, J., Shrestha, P., Pal Singh, S., 2012. Modification of seed oil composition in *Arabidopsis* by artificial microRNA-mediated gene silencing. Front. Plant Sci. 31;3:168 Bonaventure, G., Salas, J.J., Pollard, M.R., Ohlrogge, J.B., 2003. Disruption of the *FATB* gene in Arabidopsis demonstrates an essential role of saturated fatty acids in plant growth. Plant Cell. 15, 1020-1033.

Bouby, L., 1998. Two early finds of gold-of-pleasure (Camelina sp.) in middle Neolithic and Chalcolithic sites in western France. Antiquity. 72, 391-398.

Bramm, A., Dambroth, M., Schulte-Korne, S., 1990. Analysis of yield components of linseed, false flax, and poppy. Landbauforschung Volkenrode. 40, 107-114.

Ciubota-Rosie, C., Ruiz, J.R., Ramos, M.J. and Pérez, A., 2013. Biodiesel from *Camelina sativa*: a comprehensive characterisation. Fuel. 105, 572-577.

Collins-Silva, J.E., Lu, C., Cahoon, E.B, 2011. Camelina: a designer biotech oilseed crop. Inform. 22, 610-613.

Dörmann, P., Voelker, T.A., Ohlrogge, J.B., 1995. Cloning and expression in Escherichia coli of a novel thioesterase from Arabidopsis thaliana specific for long-chain acyl–acyl carrier proteins. Arch. Biochem. Biophys. 316, 612-618.

Facciotti, M.T., Yuan, L., 1998. Molecular dissection of the plant acyl-acyl carrier protein thioesterases. Fett Lipid. 100, 167-172.

Fröhlich, A., Rice, B., 2005. Evaluation of *Camelina sativa* oil as a feedstock for biodiesel production. Ind. Crop Prod. 21, 25-31.

Garcés, R., Mancha, M., 1993. One-step lipid extraction and fatty acid methyl esters preparation from fresh plant tissues. Anal. Biochem. 211, 139-143.

Gehringer, A., Friedt, W., Lühs, W., Snowdon, R.J., 2006. Genetic mapping of agronomic traits in false flax (*Camelina sativa* subsp. sativa). Genome. 49, 1555-1563.

Ghosh, S., Sen, S.K., 2007. Characterization and cloning of a stearoyl/oleoyl specific fatty acyl–acyl carrier protein thioesterase from the seeds of *Madhuca longifolia* (latifolia). Plant Physiol. Biochem. 45, 887-897.

Haas, J.A., M.A. Frederick, and B.G. Fox., 2000. Chemical and Posttranslational Modification of *Escherichia coli* Acyl Carrier Protein for Preparation of Dansyl-Acyl Carrier Proteins. *Protein Expression and Purification.* 20, 274-284.

Hawkins, D.J., Kridl, J.C., 1998. Characterization of acyl-ACP thioesterases of mangosteen (*Garcinia mangostana*) seed and high levels of stearate production in transgenic canola. Plant J. 13, 743-752.

He, X., Chen, G.Q., Lin, J.T., McKeon, T.A., 2004. Regulation of diacylglycerolacyltransferase in developing seeds of castor. Lipids 39, 865–871.

Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., Nakai, K., 2007. WoLF PSORT: protein localization predictor. Nucl. Acids Res. 35, 585-587.

Hutcheon, C., Ditt, R.F., Beilstein, M., Comai, L., Schroeder J., Goldstein, E., Shewmaker, C.K., Nguyen, T., Rocher J.D., Kiser, J., 2010. Polyploid genome of *Camelina sativa revealed by isolation of fatty acid synthesis genes. BMC Plant Biology. 10:233*

Huynh, T.T., Pirtle, R.M., Chapman, K.D., 2002. Expression of a Gossypium hirsutum cDNA encoding a FatB palmitoyl-acyl carrier protein thioesterase in Escherichia coli. Plant Physiol. Biochem. 40, 1-9

Jha, J.K., Maiti, M.K., Bhattacharjee, A., Basu, A., Sen, P.C., Sen, S.K., 2006. Cloning and functional expression of an acyl-ACP thioesterase FatB type from *Diploknema* (Madhuca) *butyracea* seeds in *Escherichia coli*. Plant Physiol. Biochem. 44, 645-655.

Jones A., Davies H.M., Voelker T.A., 1995. Palmitoyl-acyl carrier protein (ACP) thioesterase and the evolutionary origin of plant acyl-ACP thioesterases. Plant Cell. 7, 359-371

Joshi, C.P., Zhou, H., Huang, X., Chiang, V.L., 1997. Context sequences of translation initiation codon in plants. Plant Mol. Biol. 35, 993-1001.

Knutzon, D.S., Bleibaum, J.L., Nelsen, J., Kridl, J.C., Thompson, G.A., 1992. Isolation and characterization of two safflower oleoyl–acyl carrier protein thioesterase cDNA clones. Plant Physiol. 100, 1751-1758.

Koo, A.J., Ohlrogge, J.B., Pollard, M., 2004. On the export of fatty acids from the chloroplast. J. Biol. Chem. 279, 16101-16110.

Kozak, M., 1991. An analysis of vertebrate mRNA sequences: intimations of translational control. J. Cell Biol. 115, 887-903.

Lambalot, R.H., Walsh, C.T., 1997. Holo-[Acyl-Carrier-Protein] Synthase of Escherichi coli. Meth. Enzym. Mol. 27, 254-262.

Lennen, R.M., Pfleger, B.F., 2012. Engineering Escherichia coli to synthesize free fatty acids. Trends Biotechnol. 30, 659-667.

Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2_DDCT method. Methods. 25, 402-408.

Lovett, J.V., Jackson, H.F., 1980. Allelopathic activity of *Camelina sativa* (L.) Crantz inrelation to its phyllosphere bacteria. New Phytol. 86, 273-277.

Lu, C. and Kang, J., 2008. Generation of transgenic plants of a potential oilseed crop Camelina sativa by Agrobacterium-mediated transformation. Plant Cell Rep. 27, 273-278.

Martínez-Force, E., Cantisán, S., Serrano-Vega, M.J., Garcés, R., 2000. Acyl–acyl carrier protein thioesterase activity from sunflower (*Helianthus annuus* L.) seeds. Planta. 211, 673-678.

Mayer K.M., Shanklin J, 2007. Identification of amino acid residues involved in substrate specificity of plant acyl-ACP thioesterases using a bioinformatics guided approach. BMC Plant Biology. 7:1

Moreno-Pérez, A.J., Sánchez-García, A., Salas, J.J, Garcés, R. and Martínez-Force, E., 2011. Acyl-ACP thioesterases from macadamia (*Macadamia tetraphylla*) nuts: cloning, characterization and their impact on oil composition. Plant Physiol. Biochem. 49, 82-87.

Moreno-Pérez, A.J., Venegas-Calerón, M., Vaistij, F.E., Salas, J.J., Larson, T.R., Garcés, R., Graham, I.A., Martínez-Force, E., 2012. Reduced expression of FatA thioesterases in Arabidopsis affects the oil content and fatty acid composition of the seeds. Planta. 235:629-639.

Ohlrogge, J., Pollard, M., Bao, X., Focke, M., Girke, T., Ruuska, S., Mekhedov, S., Benning, C., 2000. Fatty acid synthesis: from CO2 to functional genomics. Biochem. Soc. Trans. 28, 567-573.

Pathak, M.K., Bhattacharjee, A., Ghosh, D., Ghosh, S., 2004. Acyl–acyl carrier protein (ACP)–thioesterase from developing seeds of *Brassica campestris* cv. B-54 (Agrani). Plant Sci. 166, 191-198.

Pidkowitch M.S., Nguyen H.T., Heilmann I., Ischebeck T., Shanklin J., 2007. Modulating seed ß-ketoacyl-acyl carrier protein synthase II level converts the composition of a temperate seed oil to that of a palm-like tropical oil. Proc Natl Acad Sci USA. 104: 4742-4747.

Putnam, D., Budin, J., Field, L. and Breene, W., 1993. Camelina: a promising low-input oilseed. New Crops (Janick, J. and Simon, J.E., eds). 314-322.

Rock C.O., Garwin J.L., 1979. Preparative enzymatic synthesis and hydrophobic chromatography of acyl-acyl carrier protein. J. Biol. Chem. 254, 7123-7128.

Rodríguez-Rodríguez M.F., Sánchez-García A., Salas J.J., Garcés R, Martínez-Force E., 2013. Characterization of the morphological changes and fatty acid profile of developing *Camelina sativa* seeds. Industrial Crops and Products. 50, 673-679

Salas, J.J., Ohlrogge, J.B., 2002. Characterization of substrate specificity of plant FatA and FatB acyl-ACP thioesterases. Arch. Biochem. Biophys. 403, 25-34.

Sánchez-García, A, Moreno-Pérez, A.M. Muro-Pastor, A, Salas, J.J, Garcés, R, Martínez-Force, E., 2010. Acyl-ACP thioesterases from castor (*Ricinus communis* L.): An enzymatic system appropriate for high rates of oil synthesis and accumulation. Phytochemistry. 71, 860–869.

Serrano-Vega, M.J; Garcés, R; Martínez-Forces, E, 2005. Cloning, characterization and structural model of a FatA-type thioesterase from sunflower seeds (Helianthus annuus L.). Planta. 221, 868–880Shanklin, J; Cahoon, E.B., 1998. Desaturation and related modifications of fatty acids. Annu Rev Plant Physiol Plant Mol Biol. 49, 611-641

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

Voelker T.A., Jones A., Cranmer A.M., Davies H.M., Knutzon D.S., 1997. Broad-range and binary-range acyl-acyl-carrier-protein thioesterases suggest an alternative mechanism for mediumcha in production in seeds. Plant Physiol 114, 669–677

Zurb, J., 1997. Oil-seed crop: Camelina sativa. Industrial Crops and Products, Volume 6, Issue 2, May 1997, 113-119

Table 1 **Degenerate and non-degenerate oligonucleotide primers used in this work**

Primer name	Sequence 5´- 3´
***Cs*FatA**	
Deg_FatA_F	GARATHTAYARRTAYCCNGC
Deg_FatA_R	CAYTCNCKNCKRTANTC
CsFatA_F1	TGATGTGGTTGAGATTGAGACTTGG
CsFatA_F2	GGTGATGATGAACCAAGACACAAGG
CsFatA_R1	TCTCAAGGACCCACCCAATGTAGG
CsFatA_R2	TAAGCCCAAGCATAGAATACTGAGC
CsFatASphI_F3[a]	TAGCAGCATGCATGTTGAAGCTTTCATGTAATG
CsFatASphI_F4	CTAGCATGCGTATCTGCTGATCAAGGAAGTG
CsFatAXmaI_R3	ACTCCCGGGTCATCTTGAGGACTTCTTTCTCC
CsFatA_INTR_F1	ACCATTGAAACCATCGCTAATC
CsFatA_INTR_R1	GTAATGACCTGAAGTTCGTGC
CsFatA_qPCR_F	CCAGCATGTCAATAATGTCACC
CsFatA_qPCR_R	CCTGTGTGCCAGACGTGG
***Cs*FatB**	
Deg_FatB_F	AGTGGSTGATGCTYRABTGG
Deg_FatB_R	TNACRTGYTGRTTRAYRTC
CsFatB_F1	AGGCGTTCTGACATGCTTATG
CsFatB_F2	CATGTTAAGACTGCCGGACTGC
CsFatB_R1	TGCTGGTTAATATCAAGGTCACTT
CsFatB_R2	AATCCCGAACTAGCCAATCACG
CsFatB_F3	ATGGTGGCCACCTCTGC
CsFatB_R3	TTACGGTGTAGTTCCCCAAGTT
CsFatBSphI_F4	TAGCAGCATGCCCCAGGCGTTCTGACATG
CsFatBXmaI_R4	TATCCCGGGTTACGGTGTAGTTCCCCAAG
CsFatB_INTR_F1	CGTATGAAATAGGTGCTGATCG
CsFatB_INTR_R1	CTTTACATTATTCACATGCTGG
CsFatB_qPCR_F	CTTGATGACAAGACTGCTGACT
CsFatB_qPCR_R	GTGAGCGACTGAAGGACACT
CsActin	
CsActin_qPCR_F	TTGGAAGGATCTGTACGGTAAC
CsActin_qPCR_F	TGTGAACGATTCCTGGACC

[a] The restriction sites included are underlined.
H: A or C or T; K: G or T; N: A or C or G or T; R: A or G; Y: C or T.
All primers were synthesized by Eurofins MWG Operon (Germany).

Table 2
Exon and intron lengths (bp) in FatA from *Camelina sativa* and *Arabidopsis thaliana*

Gene	E1	iI	E2	iII	E3	iIII	E4	iIV	E5	iV	E6	iVI	E7
CsFatA1	358	108	134	84	113	169	96	106	88	82	69	89	246
CsFatA2	361	103	134	86	113	171	96	105	88	82	69	88	249
CsFatA3	361	103	134	85	113	260	96	105	88	82	69	88	249
AtFatA1	329	100	134	104	114	144	96	78	88	84	69	86	256
AtFatA2	330	121	134	73	114	157	96	74	157	88	273		

Exon and intron lengths (bp) in FatB from *Camelina sativa* and *Arabidopsis thaliana*

Gene	E1	iI	E2	iII	E3	iIII	E4	iIV	E5
CsFatB1	511	92	134	221	111	139	174	87	321
CsFatB2	511	91	134	305	111	141	174	89	321
CsFatB3	511	90	134	314	111	131	174	90	321
AtFatB	499	81	135	227	113	227	172	90	321

Table 3

Fatty acid composition of *E.Coli* cells containing recombinant plasmid. The date are the averages of independent samples.

	16:0	16:1ω7	17:0Δ[a]	18:0	18:1ω9	19:0Δ[b]	UFA[c]	STA[d]	UFA/STA
pQE80	52.6 ± 0.6	9.9 ± 0.5	21.3 ± 0.4	1.8 ± 0.4	11.2 ± 0.2	3.2 ± 0.1	45.58 ± 0.97	54.42 ± 0.97	0.84 ± 0.15
pQE80_CsFatA	55.9 ± 1.3	21.7 ± 0.7	13.1 ± 1.3	3.7 ± 0.5	5.2 ± 0.1	0.5 ± 0.1	40.36 ± 0.73	59.64 ± 0.73	0.68 ± 0.01
pQE80_CsFatB	43.1 ± 1.8	29.3 ± 0.7	7.1 ± 0.9	7.0 ± 1.0	12.5 ± 0.7	1.1 ± 0.1	49.89 ± 0.74	50.11 ± 0.94	1.00 ± 0.0

a cis-9,10-Methylen-hexadecanoic acid, cyclopropane derivative from 16:1.

b cis-11,12-Methylen-octadecanoic acid, cyclopropane derivative from 18:1.

c Unsaturated fatty acids and derivatives; 16:1ω7 + 17:0Δ + 18:1ω9 + 19:0Δ.

d Saturated fatty acids; 16:0 + 18:0.

Table 4:
Kinetic parameters of the purified recombinant CsFatA and CsFatB proteins acting on different acyl-ACP substrate

	Substrate	Km (μM)	Vmax (nkat/mg prot)	Kcat (s^{-1})	Kcat/Km (s^{-1} μM^{-1})
CsFatA	**16:0-ACP**	1.82	4.00	0.33	0.18
	16:1-ACP	1.90	2.12	0.17	0.09
	18:0-ACP	0.62	5.10	0.42	0.68
	18:1-ACP	4.40	92.56	7.61	1.73
CsFatB	**16:0-ACP**	6.68	57.45	4.51	0.67
	16:1-ACP	1.72	0.70	0.05	0.03
	18:0-ACP	2.84	6.01	0.47	0.17
	18:1-ACP	2.06	9.95	0.78	0.38

Data represent the mean of 3 independent determinations; SD < 10% of mean value

Figure Legends

Figure 1. Alignment of the deduced amino acid sequences of acyl-ACP thioesterase A enzymes from *Camelina sativa* (*Cs*FatA1, AFQ60947.1; *Cs*FatA2, AFQ60948.1; *Cs*FatA3, AFQ60946.1), *Arabidopsis thaliana* (*At*FatA1, NP_189147.1; *At*FatA2, NP_193041.1) and *Zea mays* (*Zm*FatA, DAA40472.1). Identical amino acids are shaded in black, whereas conserved residues are shaded in grey. The amino acids considered to constitute the signal peptide are boxed. The three conserved residues that constitute the catalytic triad are indicated with an asterisk (Asn-273; His-275; Gln-311), and the residues involved in specific substrate recognition and related with the thioesterase activity are indicated by the arrowheads (Gly-135; Ala-137; Arg-143; Lys-144; Thr-182; Arg-184; Arg-212; Arg-213; Lys-216). The conservative changes in the amino acid sequence between the three CsFatA alleles are marked by closed circles and the semi-conservative changes between them by open circles.

Figure 2. Alignment of the deduced amino acid sequences of the acyl-ACP thioesterase B enzymes from *Camelina sativa* (*Cs*FatB1, AFQ60949.1; *Cs*FatB2, AFQ60950.1; *Cs*FatB3, AFQ60951.1), *Arabidopsis thaliana* (*At*FatB, CAA85388.1) and *Zea mays* (*Zm*FatB, AFW85914.1). Identical amino acids are shaded in black, whereas conserved residues are shaded in grey. The amino acids considered to constitute the signal peptide are boxed and the hydrophobic region in FatB is underlined. The three conserved residues that constitute the catalytic triad are indicated by an asterisk (Asn-319; His-321; Cys-383), and the residues involved in the specific substrate recognition and related with the thioesterase activity are indicated by the arrowheads (Asp-186; Gly-189; Met-233; Arg-235; Lys-267; Glu-270). The conservative changes in the amino acid sequence between the three CsFatB alleles are indicated by open circles.

Figure 3. Phylogenetic comparison of plant acyl-ACP thioesterase enzymes. Plant species included in the phylogenetic tree are: Amborellaceae (*Amborella trichopoda*), Apiaceae (*Coriandrum sativum*), Asteraceae (*Carthamus tinctorius*, *Helianthus annuus*), Brassicaceae (*Arabidopsis thaliana*, *A. lyrata*, *Camelina sativa*, *Capsella rubela, Brassica juncea*, *Brassica napus*, *Brassica rapa*), Briophyta (*Physcomitrella patens*), Clusiaceae (*Garcinia mangostana*), Coniferophyta (*Picea sitchensis*), Cucurbitacea (*Cucumis sativus*), Euphorbiaceae (*Jatropha curcas*, *Ricinus communis*),

Fabaceae (*Arachis hypogaea*, *Cicer arietinum*, *Glycine max*, *Medicago truncatula*), Iridaceae (*Iris germánica*, *I. tectorum*), Lamiaceae (*Perilla frutescens*), Lycopodiophyta (*Selaginella moellendorffii*), Lythraceae (*Cuphea sps*), Malvaceae (*Gossypium hirsutum*, *Theobroma cacao*), Myristicaceae (*Myristica fragans*), Poaceae (*Aegilops tauschii*, *Brachypodium distachyon*, *Cocos nucifera*, *Elaeis guineensis*, *Hordeum vulgare*, *Oryza sativa*, *Setaria italica*, *Sorghum bicolor*, *Zea mays*), Proteaceae (*Macadamia tetraphylla*), Rosaceae (*Fragaria vesca*, *Prunus persica*), Rutaceae (*Citrus sinensis*), Salicaceae (*Populus trichocarpa*), Solanaceae (*Capsicum annuum*, *C. chinense*, C. *frutescens*, *Nicotiana tabacum*, *Solanum lycopersicum*), Theaceae (*Camellia oleífera*), and Vitaceae (*Vitis vinifera*). The phylogenetic tree was rooted using the sequence of acyl-ACP thioesterase from a bacterium of the genus *Desulfovibrio*.

Figure 4. Coomassie blue stained SDS-PAGE showing recombinant *Camelina sativa* acyl-ACP thioesterases A (panel A) and acyl-ACP thioesterase B (panel B). Lane 1, soluble fraction; lane 2, soluble fraction not retained on the Ni-NTA Agarose column (Qiagen); lane 3, Ni-NTA Agarose wash; lane 4, acyl-ACP thioesterase purification.

Figure 5. Substrate specificity of *Camelina sativa* acyl-ACP thioesterases expressed in *E. coli.* The activity was measured with the purified His-tagged *Cs*FatA (black columns) and His-tagged *Cs*FatB (white columns) enzymes, testing different acyl-ACP substrates. The data represent the mean (±SD) from three independent assays.

Figure 6. Expression of the *CsFatA* (black columns) and *CsFatB* (white columns) genes in vegetative tissues and developing seeds from *Camelina sativa* determined by QRT-PCR. The data represent the mean values ± SD of three independent assays.

Signal Peptide

```
CsFatA1  MLKLSCNVTDS-KLQRSLLFFSHSHRSDPVNFTRRRIVSSSVSCSQTKKTGLVDPLRAVVSADQG--SVIRAEQGLGT-L  76
CsFatA2  MLKLSCNVTDSSKLQRSLLFFSHSHRSDPVNCIRRRIVASSVSCSQTKKTGLVDPLRAVVSADQG--SVIRAEQGLGT-L  77
CsFatA3  MLKLSCNVTDSSKLQRSLLFFSHSHRSDPVNCIRRRIVSSSVSCSQTKKTGLVDPLRAVVSADQG--SVIRAEQGLGT-L  77
AtFatA1  MLKLSCNVTDS-KLQRSLLFFSHSYRSDPVNFIRRRIVS----CSQTKKTGLV-PLRAVVSADQG--SVV---QGLAT-L  68
AtFatA2  MLKLSCNVTDH--IHN---LFSNSRRIFVPVHRQTRPIS----CFQLKK----EPLRAILSADHGNSSVRVADTVSGTSP  67
ZmFatA   --MLRCHAPPQ--CGRAPLRHHGRWESSPAPGVVVRCTRG---APQVSGIEAASPDHAAATAVAA----KAEGGDARPSL  69

CsFatA1  ADQLRLGSLTEDGLSYKEKFVVRSYEVGSNKTATIETIANLLQEVGCNHAQSVGFSTDGFATTPTMRKLHLIWVTARMHI  156
CsFatA2  ADQLRLGSLTEDGLSYKEKFVVRSYEVGSNKTATIETIANLLQEVGCNHAQGVGFSTDGFATTPTMRKLHLIWVTARMHI  157
CsFatA3  ADQLRLGSLTEDGLSYKEKFVVRSYEVGSNKTATIETIANLLQEVGCNHAQSVGFSTDGFATTPTMRKLHLIWVTARMHI  157
AtFatA1  ADQLRLGSLTEDGLSYKEKFVVRSYEVGSNKTATVETIANLLQEVGCNHAQSVGFSTDGFATTTTMRKLHLIWVTARMHI  148
AtFatA2  ADRLRFGRLMEDGFSYKEKFIVRSYEVGINKTATIETIANLLQEVACNHVQNVGFSTDGFATTLTMRKLHLIWVTARMHI  147
ZmFatA   AERLRLGSLLEDGLSYKESFIVRCYEVGINKTATVETIANLLQEVGCNHAQSVGFSTDGFATTTTMRKLGLIWVTNRMHI  149

CsFatA1  EIYNYPAWGDVVEIETWCQSEGRIGTRRDWILKDCATGEVTGRATSKWVMMNQDTRRLQKVSDDVRDEYLVFCPQEPRLA  236
CsFatA2  EIYNYPAWGDVVEIETWCQSEGRIGTRRDWILKDCATGEVTGRATSKWVMMNQDTRRLQKVSDDVRDEYLVFCPQEPRLA  237
CsFatA3  EIYNYPAWGNVVEIETWCQSEGRIGTRRDWILKDCATGEVTGRATSKWVMMNQDTRRLQKVSDDVRDEYLVFCPQEPRLA  237
AtFatA1  EIYKYPAWGDVVEIETWCQSEGRIGTRRDWILKDSVTGEVTGRATSKWVMMNQDTRRLQKVSDDVRDEYLVFCPQEPRLA  228
AtFatA2  EIYKYPAWSDVVEIETWCQSEGRIGTRRDWILKDCATGEVIGRATSKWVMMNQDTRRLQRVTDEVRDEYLVFCPPEPRLA  227
ZmFatA   EIYKYPAWGDVVEIETWCQEDGKIGTRRDWILKDLSTGEVTGRATSKWVMMNQNTRRLQRVSDDVRDEVFIHCPKTPRLA  229

CsFatA1  FPEENNRSLKKIPKLEDPAQYSMIGLKPRRADLDMNQHVNNVTYIGWVLESVPQEIVDTHELQVITLDYRRECQQDDVVD  316
CsFatA2  FPEENNRSLKKIPKLEDPAQYSMIGLKPRRADLDMNQHVNNVTYIGWVLESVPQEIVHTHELQVITLDYRRECQQDDVVD  317
CsFatA3  FPEENNRSLKKIPKLEDPAQYSMIGLKPRRADLDMNQHVNNVTYIGWVLESVPQEIVDTHELQVITLDYRRECQQDDVVD  317
AtFatA1  FPEENNRSLKKIPKLEDPAQYSMIGLKPRRADLDMNQHVNNVTYIGWVLESIPQEIVDTHELQVITLDYRRECQQDDVVD  308
AtFatA2  FPEENNSSLKKIPKLEDPAQYSMLGLKPRRADLDMNQHVNNVTYIGWVLESIPQEIIDTHELKVITLDYRRECQQDDIVD  307
ZmFatA   FPEENNGSLKKIPNLSDPAQYSRLGIVPRRADLDMNQHVNNVTYIGWVLESIPQDIIDTHELQTITLDYRRECQHDDIVD  309

CsFatA1  SLTT--------SEIGGTNGSATSGTQGHNDSQFLHLLRLSGDGQEINRGTTLWRKKSSR  368
CsFatA2  SLTT--------SEIGGTNGSATSGTQGHNDSQFLHLLRLSGDGQEINRGTTLWRKKSSR  369
CsFatA3  SLTT--------SEIGGTNGSATSGTQGHNDSQFLHLLRLSGDGQEINRGTTLWRKKSSR  369
AtFatA1  SLTTTT------SEIGGTNGSATSGTQGHNDSQFLHLLRLSGDGQEINRGTTLWRKKPSS  362
AtFatA2  SLTTSETPNEVVSKLTGTNGSTTSSKREHNESHFLHILRLSENGQEINRGRTQWRKKSSR  367
ZmFatA   SLTYVEE-----GEERSMNGSASSVPHTEQRRQFLHCLRFAANGDEINRGRTVWRKLAR-  363
```

Figure 1

Signal Peptide

Hydrophobic Region

```
CsFatB1  MVATSATSSFFPVPSSSSSLDPNGKGG-NKIGSTNF--AGLNSPPNS--ARMKVKPNAQAPPKINGKKVGLPGSVDIVRT 75
CsFatB2  MVATSATSSFFPVPSSSSSLDPIGKGG-NKIGSTNF--AGLNSPPNS--GRMKVKPNAQAPPKINGKKVGLPGSVDIVRS 75
CsFatB3  MVATSATSSFFPVPSSSSS-LDPNGKGGGNKIGSTNF--AGLNSPPNS--GRMKVKPNAQAPPKINGKKVGLPGSVDIVRT 75
AtFatB   MVATSATSSFFPVP--SSSLDPNGKGN--KIGSTNL--AGLNSTPNS--GRMKVKPNAQAPPKINGKKVGLPGSVDIVRT 72
ZmFatB   MAGSLAASAFFPGPGASPAASAKTSKNLAGELPDNLSVRGIVATPNAPSGNMQVKAQAQALPKVNGTKVNLKN----ASP 76

CsFatB1  DTETSSHPPAPRTFINQLPDWSMLLAAITTIFLAAEKQWMMLDWKPRRSDMLMDPFGIGRIVQDGLVFRQNFSIRSYEIG 155
CsFatB2  DTETSSHPPAPRTFINQLPDWSMLLAAITTIFLAAEKQWMMLDWKPRRSDMLMDPFGLGRIVQDGLVFRQNFSIRSYEIG 155
CsFatB3  DTETSSHPPAPRTFINQLPDWSMLLAAITTIFLAAEKQWMMLDWKPRRSDMLMDPFGIGRIVQDGLVFRQNFSIRSYEIG 155
AtFatB   DTETSSHP-APRTFINQLPDWSMLLAAITTIFLAAEKQWMMLDWKPRRSDMLVDPFGIGRIVQDGLVFRQNFSIRSYEIG 151
ZmFatB   DTEEAIPYIAPKTFYNQLPDWSMLLAAVTTIFLAAEKQLTLLDWKPKKPDMLVDTFGFGRIIQDGLVFRQNFWIRSYEIG 156

CsFatB1  ADRAASIETVMNHLQETALNHVKTAGLLGDGFGSTPEMFKKNLIWVVTRMQVVVDKYPTWGDVVEVDTWVSQSGKNGMRR 235
CsFatB2  ADRAASIETVMNHLQETALNHVKTAGLLGDGFGSTPEMFKKNLIWVVTRMQVVVDKYPTWGDVVEVDTWVSQSGKNGMRR 235
CsFatB3  ADRAASIETVMNHLQETALNHVKTAGLLGDGFGSTPEMFKKNLIWVVTRMQVVVDKYPTWGDVVEVDTWVSQSGKNGMRR 235
AtFatB   ADRSASIETVMNHLQETALNHVKTAGLLGDGFGSTPEMFKKNLIWVVTRMQVVVDKYPTWGDVVEVDTWVSQSGKNGMRR 231
ZmFatB   ADRTASIETLMNHLQETALNHVKTAGLLGDGFGATPEMSKRNLIWVVSKIQLLVEQYPSWGDTVQVDTWVAAAGKNGMRR 236

CsFatB1  DWLVRDCNTGETLTRASSVWVMMNKLTRRLSKIPEEVRGEIEPYFVNSDPVLTEDSRKLTKLDDKT------ADFVRSGL 309
CsFatB2  DWLVRDCNTGETLTRASSVWVMMNKLTRRLSKIPEEVRGEIEPYFVNSDPVLTEDSRKLTKLDDKT------ADYVRSGL 309
CsFatB3  DWLVRDCNTGETLTRASSVWVMMNKLTRRLSKIPEEVRGEIEPYFVNSDPVLTEDSRKLTKLDDKT------ADYVRSGL 309
AtFatB   DWLVRDCNTGETLTRASSVWVMMNKLTRRLSKIPEEVRGEIEPYFVNSDPVLAEDSRKLTKIDDKT------ADYVRSGL 305
ZmFatB   DWHVRDENSGRTILRATSVWVMMNKNTRRLSKMPDEVRAEIGPYFNGRSAITDEQSEKLAKPGSAPDGGTMKQQFIRKGL 316

CsFatB1  TPRWSDLDVNQHVNNVKYIGWILESAPVGIMERQKLKSMTLEYRRECGRDSVLQSLTAVSGCDMGNLATAGDVECQHLLR 389
CsFatB2  TPRWSDLDVNQHVNNVKYIGWILESAPVGIMEKQKLKSMTLEYRRECGRDSVLQSLTAVSGCDMGNLATAGDVECQHLLR 389
CsFatB3  TPRWSDLDVNQHVNNVKYIGWILESAPVGIMERQKLKSMTLEYRRECGRDSVLQSLTAVAGCDMGNLATAGDVECQHLLR 389
AtFatB   TPRWSDLDVNQHVNNVKYIGWILESAPVGIMERQKLKSMTLEYRRECGRDSVLQSLTAVTGCDIGNLATAGDVECQHLLR 385
ZmFatB   TPRWGDLDVNQHVNNVKYIGWILESAPISILEQHELASMTLDYRKECGRDSVLQSLTTVAGERMDGHAGDSTIQCQHLLQ 396

CsFatB1  LQDGAEVVRGRTEWSSKTP---TTTWGTTP-------- 416
CsFatB2  LQDGAEVVRGRTEWSSKTP---TTTWGTTP-------- 416
CsFatB3  LQDGAEVVRGRTEWSSKTP---TTTWGTTP-------- 416
AtFatB   LQDGAEVVRGRTEWSSKTP---TTTWGTAP-------- 412
ZmFatB   LESGADIVKAHTEWRPKRARGEGSLRGASGVFRPRMHE 434
```

Figure 2

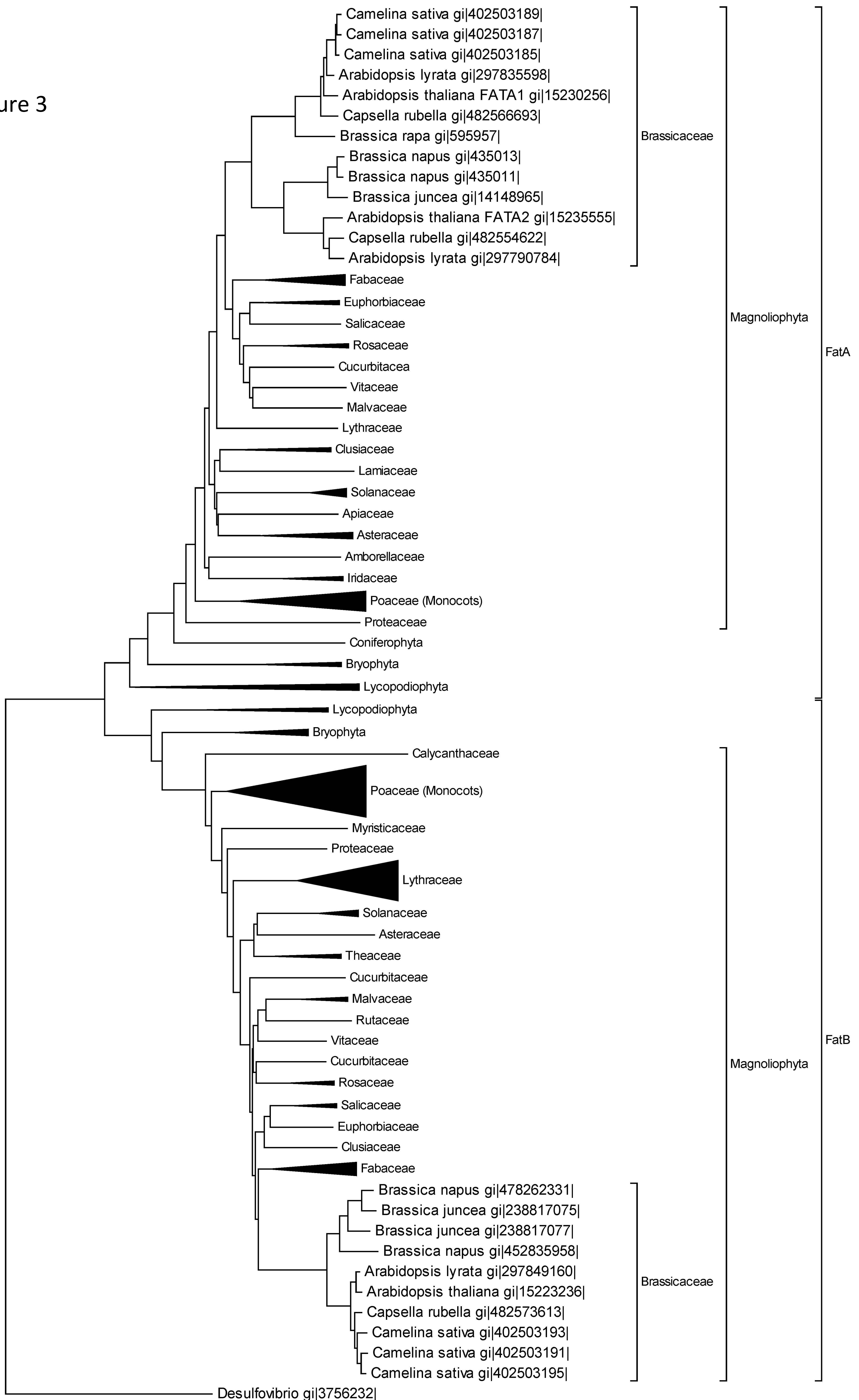
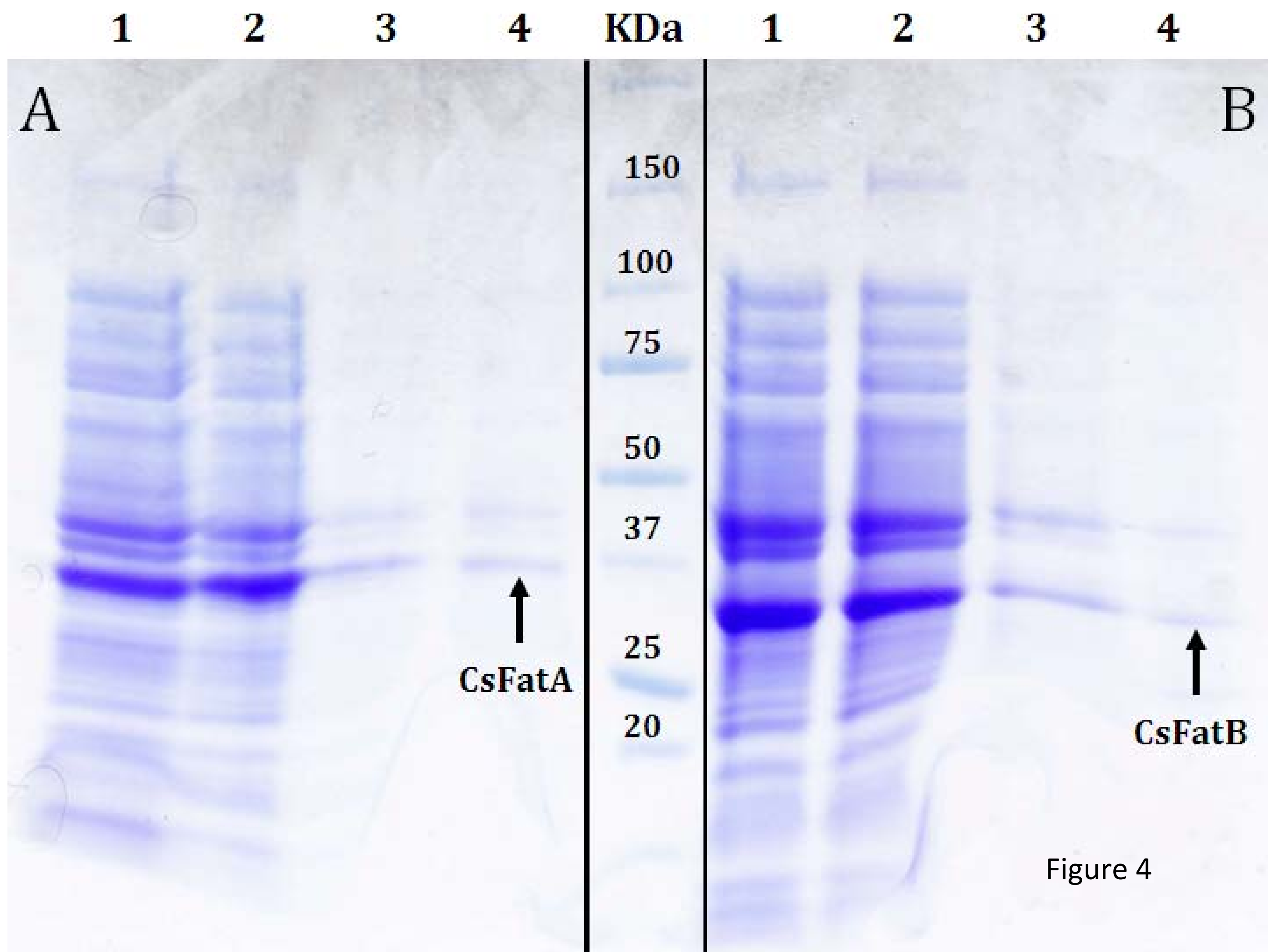


Figure 3
Camelina sativa gi|402503189|
Camelina sativa gi|402503187|
Camelina sativa gi|402503185|
Arabidopsis lyrata gi|297835598|
Arabidopsis thaliana FATA1 gi|15230256|
Capsella rubella gi|482566693|
Brassica rapa gi|595957|
Brassica napus gi|435013|
Brassica napus gi|435011|
Brassica juncea gi|14148965|
Arabidopsis thaliana FATA2 gi|15235555|
Capsella rubella gi|482554622|
Arabidopsis lyrata gi|297790784|
Brassicaceae
Fabaceae
Euphorbiaceae
Salicaceae
Rosaceae
Cucurbitacea
Vitaceae
Malvaceae
Lythraceae
Clusiaceae
Lamiaceae
Solanaceae
Apiaceae
Asteraceae
Amborellaceae
Iridaceae
Poaceae (Monocots)
Proteaceae
Coniferophyta
Bryophyta
Lycopodiophyta
Magnoliophyta
FatA
Lycopodiophyta
Bryophyta
Calycanthaceae
Poaceae (Monocots)
Myristicaceae
Proteaceae
Lythraceae
Solanaceae
Asteraceae
Theaceae
Cucurbitaceae
Malvaceae
Rutaceae
Vitaceae
Cucurbitaceae
Rosaceae
Salicaceae
Euphorbiaceae
Clusiaceae
Fabaceae
Brassica napus gi|478262331|
Brassica juncea gi|238817075|
Brassica juncea gi|238817077|
Brassica napus gi|452835958|
Arabidopsis lyrata gi|297849160|
Arabidopsis thaliana gi|15223236|
Capsella rubella gi|482573613|
Camelina sativa gi|402503193|
Camelina sativa gi|402503191|
Camelina sativa gi|402503195|
Brassicaceae
Magnoliophyta
FatB
Desulfovibrio gi|3756232|
0.05


1
2
3
4
KDa
1
2
3
4
A
B
150
100
75
50
37
25
20
CsFatA
CsFatB
Figure 4

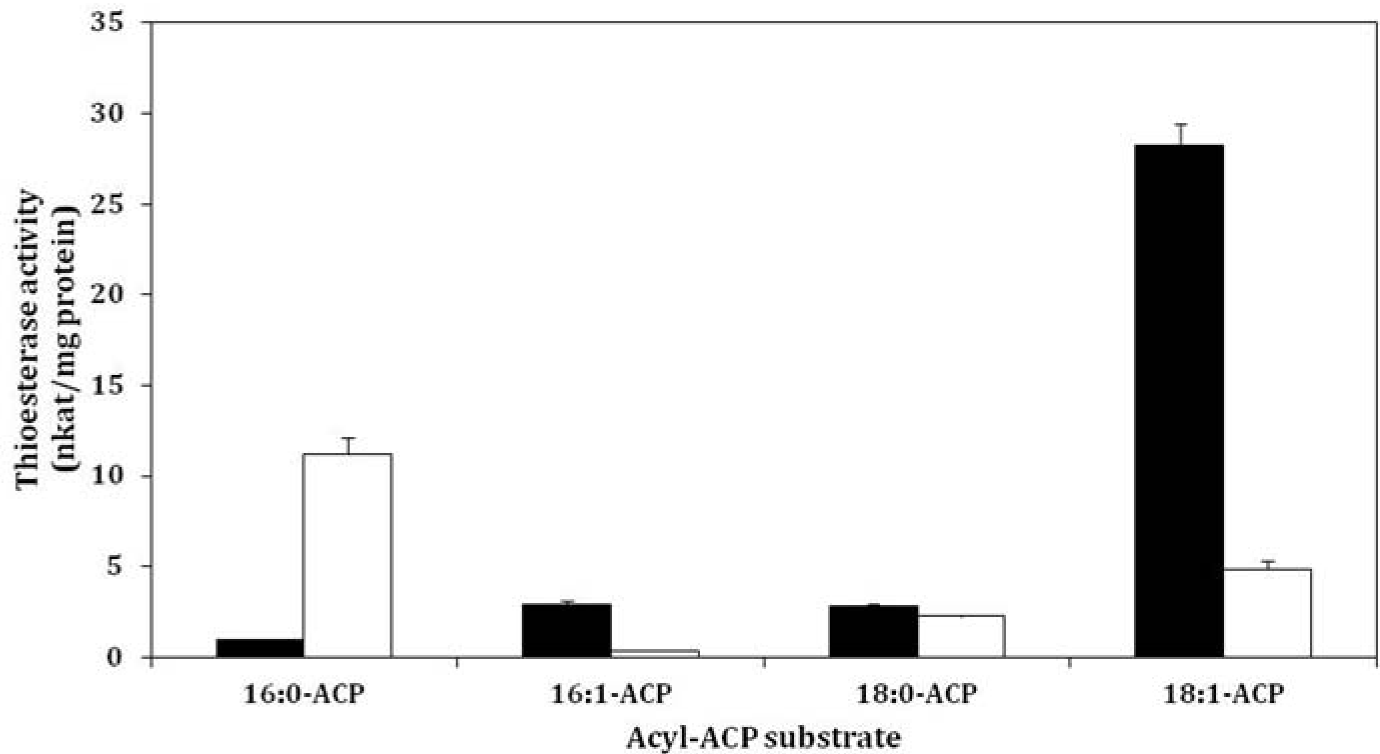


Figure 5

Figure 6