Plant Biotechnology Journal

Plant Biotechnology Journal (2014), pp. 1–11 doi: 10.1111/pbi.12167

Targeted mutation of Δ12 and Δ15 desaturase genes in hemp produce major alterations in seed fatty acid composition including a high oleic hemp oil

Monika Bielecka[1,†,§], Filip Kaminski[1,§], Ian Adams[1,‡], Helen Poulson[1], Raymond Sloan[2], Yi Li[1], Tony R. Larson[1], Thilo Winzer[1] and Ian A. Graham[1,*]

[1]*Centre for Novel Agricultural Products, Department of Biology, University of York, York, UK*
[2]*Biorenewables Development Centre, The Biocentre, York Science Park, Heslington, York, UK*

Received 25 October 2013;
revised 16 December 2013;
accepted 22 December 2013.
**Correspondence* (fax 00 44 (0) 1904 328762; Tel 00 44 (0) 1904 328750; email ian.graham@york.ac.uk)
†Present address: Department of Biology and Pharmaceutical Botany, Faculty of Pharmacy, Medical University of Wroclaw, Borowska 211, 50-556 Wroclaw, Poland.
‡Present address: Crop and Food Security Programme, Food and Environment Research Agency, York, UK
§These authors contributed equally to this work.

Keywords: *Cannabis sativa*, TILLING, reverse genetics, metabolic engineering, genome mining, transcriptomics.

Summary

We used expressed sequence tag library and whole genome sequence mining to identify a suite of putative desaturase genes representing the four main activities required for production of polyunsaturated fatty acids in hemp seed oil. Phylogenetic-based classification and developing seed transcriptome analysis informed selection for further analysis of one of seven Δ12 desaturases and one of three Δ15 desaturases that we designate *CSFAD2A* and *CSFAD3A*, respectively. Heterologous expression of corresponding cDNAs in *Saccharomyces cerevisiae* showed CSFAD2A to have Δ*x*+3 activity, while CSFAD3A activity was exclusively at the Δ15 position. TILLING of an ethyl methane sulphonate mutagenized population identified multiple alleles including non-sense mutations in both genes and fatty acid composition of seed oil confirmed these to be the major Δ12 and Δ15 desaturases in developing hemp seed. Following four backcrosses and sibling crosses to achieve homozygosity, *csfad2a-1* was grown in the field and found to produce a 70 molar per cent high oleic acid ($18:1^{\Delta9}$) oil at yields similar to wild type. Cold-pressed high oleic oil produced fewer volatiles and had a sevenfold increase in shelf life compared to wild type. Two low abundance octadecadienoic acids, $18:2^{\Delta6,9}$ and $18:2^{\Delta9,15}$, were identified in the high oleic oil, and their presence suggests remaining endogenous desaturase activities utilize the increased levels of oleic acid as substrate. Consistent with this, CSFAD3A produces $18:2^{\Delta9,15}$ from endogenous $18:1^{\Delta9}$ when expressed in *S. cerevisiae*. This work lays the foundation for the development of additional novel oil varieties in this multipurpose low input crop.

Introduction

The seeds of *Cannabis sativa* L. (hemp, marijuana) have been an important source of oil and protein in human nutrition dating back to Neolithic times in ancient China (Li, 1974). *Cannabis sativa* has an annual life cycle and is mostly dioecious with male and female flowers borne on separate individuals. Selective breeding has produced marijuana strains accumulating high levels of psychoactive cannabinoids in the female flowers and hemp cultivars typically having low levels of cannabinoids but good fibre and/or seed oil traits. The draft genome of the marijuana drug strain Purple Kush and comparison of its female flower transcriptome with that of the hemp cultivar Finola demonstrated much higher expression levels of genes involved in cannabinoid production in the drug strain (van Bakel *et al.*, 2011). Furthermore, single nucleotide variant analysis revealed a distinct separation between the hemp and marijuana strains.

Finola (breeder code FIN-314) was developed in Finland as an oilseed hemp variety (Callaway and Laakkonen, 1996). It is dioecious and suitable for seed harvest by conventional agricultural machinery and can yield over 2000 kg/ha seed under good conditions (Callaway, 2004). Hemp has modest agrochemical requirements, is an excellent break crop and is suited to warm-to-temperate growing conditions (Callaway, 2004). At over 80% in polyunsaturated fatty acids (PUFAs), hemp seed oil rivals most of the commonly used vegetable oils. At 56% linoleic acid (LA, $18:2^{\Delta9,12}$) and 22% *alpha*-linolenic acid (ALA, $18:3^{\Delta9,12,15}$), hemp oil is a rich source of these essential fatty acids. In addition, hemp oil also contains *gamma*-linolenic acid (GLA, $18:3^{\Delta6,9,12}$) and stearidonic acid (SDA, $18:4^{\Delta6,9,12,15}$), which occur at about 4% and 2%, respectively, in Finola (Callaway, 2004). Here, we have used the recognized chemical nomenclature for fatty acids, which indicate the position of double bonds relative to the carboxyl group. Another frequently used terminology references the methyl end of the fatty acid with, for example, ALA ($18:3^{\Delta9,12,15}$) and SDA ($18:4^{\Delta6,9,12,15}$) being referred to as omega-3 fatty acids, whereas LA ($18:2^{\Delta9,12}$) and GLA ($18:3^{\Delta6,9,12}$) are referred to as omega-6 fatty acids.

Two multifunctional classes of desaturases have been found in plants, one soluble and the other membrane bound (Shanklin and Cahoon, 1998). In plants, C16- and C18-fatty acids are synthesized in the stroma of plastids and, following desaturation of 18:0

Please cite this article as: Bielecka, M., Kaminski, F., Adams, I., Poulson, H., Sloan, R., Li, Y., Larson, T.R., Winzer, T. and Graham, I.A. (2014) Targeted mutation of Δ12 and Δ15 desaturase genes in hemp produce major alterations in seed fatty acid composition including a high oleic hemp oil. *Plant Biotechnol. J.*, doi: 10.1111/pbi.12167

to 18:1 by a soluble Δ9 stearoyl-ACP desaturase, contribute to the assembly of complex membrane lipids (Ohlrogge and Browse, 1995). Further desaturation of fatty acids in membrane lipids of the chloroplast and endoplasmic reticulum (ER) is carried out by the membrane-bound desaturases, a number of which have been designated FAD2 TO FAD8 based on work in Arabidopsis (Ohlrogge and Browse, 1995). In the current study, we were particularly interested in the FAD2 and FAD3 enzymes because these are responsible for the Δ12 desaturation of oleic acid ($18:1^{\Delta 9}$) to LA ($18:2^{\Delta 9,12}$) and the Δ15 desaturation of LA to ALA ($18:3^{\Delta 9,12,15}$), respectively. Production of GLA ($18:3^{\Delta 6,9,12}$) from LA and SDA ($18:4^{\Delta 6,9,12,15}$) from ALA requires the action of a Δ6 desaturase which we also expect to find expressed in developing hemp seeds. Our principle objective was to perform a first characterization of the membrane-bound desaturases that determine fatty acid composition in hemp seed oil using a combination of heterologous expression and reverse genetics. The publication of the Purple Kush and Finola genomes and transcriptome sequences during the course of our work (van Bakel *et al.*, 2011) also allowed us to perform a comprehensive *in-silico* analysis of relevant fatty acid desaturases. A valuable outcome of the reverse genetic approach is that seed oil fatty acid composition is altered with the potential to improve the crop for industry. We were interested in developing a high oleic acid hemp seed oil because similar developments in crops such as soybean have opened up new markets due to increased oxidative stability (Kinney, 1998), which is a particular problem with hemp oil given the high levels of PUFAs.

Results

Genome mining reveals multiple copies of soluble and membrane-bound desaturases in *Cannabis sativa*

We isolated mRNA from the upturned (U) stage of embryo development of the Finola variety, because this represents a stage of significant storage oil deposition in the form of triacylglycerol (TAG) in dicotyledonous oilseeds (Baud *et al.*, 2002), and used this for cDNA library construction. We generated 1893 expressed sequence tags (ESTs) from the U stage cDNA library by conventional Sanger sequencing and a BLASTX similarity search revealed 11 ESTs with homology to desaturase genes. Two of the resulting unigenes contained an incomplete open reading frame (ORF) giving a predicted amino acid sequence with homology to the Δ12 desaturases. Two homologous full-length cDNA sequences were obtained by RACE PCR, and the corresponding genes were named *CSFAD2A* and *CSFAD2B* (Table S1). We also cloned a *FAD3* desaturase fragment by PCR amplification using degenerate primers (Lee *et al.*, 1998) on the U stage cDNA. RACE PCR produced a 1188-bp full-length cDNA sequence that we name *CSFAD3A* (Table S1).

We used *CSFAD2A*, *CSFAD2B*, *CSFAD3A* and other previously characterized plant membrane-bound Δ12 (FAD2), Δ15 (FAD3) and Δ6/Δ8 sphingo-lipid, as well as the soluble Δ9 stearoyl-ACP desaturases as queries to retrieve additional membrane-bound and soluble desaturase sequences from the genome sequences of Purple Kush (canSat3) and Finola (Finola1) (van Bakel *et al.*, 2011). This resulted in the identification of putative sequences for seven FAD2 (designated *CSFAD2A* to *CSFAD2G*), three FAD3 (designated *CSFAD3A*, *CSFAD3B* and *CSFAD3C*), two genes with homology to both Δ8-sphingo-lipid desaturases and Δ6 fatty acid desaturases (designated *CSD8* and *CSD6*) and five Δ9 stearoyl-ACP desaturases (designated *CSSACPD-A* to *CSSACPD-E*) in the more complete genome of the Purple Kush variety. For all but *CSFAD2F* and *CSFAD3C* orthologous sequences were also identified in the Finola genome (Table S1), which probably reflects the draft nature of this genome.

All deduced *C. sativa* amino acid sequences from both varieties were aligned with desaturase sequences from other plant species using ClustalX (Figures S1–S4). Phylogenetic trees were calculated from distance matrices by the neighbour-joining algorithm using desaturase sequences retrieved from the Purple Kush genome (Figure 1). All three subclasses of membrane-bound desaturases from *C. sativa* contain three histidine cluster motifs involved in binding the di-iron complex. CSD8 and CSD6 are homologous, but CSD8 shows greatest similarity with a number of functionally characterized Δ8 sphingo-lipid desaturases, while CSD6 is most similar to functionally characterized Δ6 fatty acid desaturases (Figures 1 and S3). In addition to the three histidine boxes, both genes contain a conserved HPGG motif within a cyt b5-like domain, the histidine residue of which is essential for enzyme activity (Sayanova *et al.*, 1999b). SACPD family members contain the EXXH motif involved in binding the di-iron complex together with additional glutamine residues involved in coordinating the di-iron complex, which are typical for this class of desaturases (Shanklin and Cahoon, 1998). Alignment of genomic and cDNA sequences revealed that all members of the *CSFAD2* and *CSD6* subclasses of microsomal desaturases consist of a single exon, while the gene arrangement of all members of the *CSFAD3* subclass contain eight exons and seven introns within the gene. This arrangement was conserved across the members of this subclass (summarized in Table S1). For the plastidial stearoyl-ACP desaturases, two (*CSSACPD-A* and *CSSACPD-B*) contain three exons and the remaining three contain two exons consistent with the phylogenetic arrangement (Table S1, Figure 1).

Deep sequencing of the developing seed transcriptome identifies candidate desaturases involved in modifying fatty acid composition of seed oil

Expressed sequence tag libraries were prepared by deep sequencing cDNA prepared from RNA isolated from torpedo (T), U and filled-not-desiccated (FND) stages of Finola embryo development as depicted in Figure 2a. Raw reads were mapped to the open reading frames of 17 putative desaturase genes as detailed in Table S1. Three of the five plastidial stearoyl-ACP desaturases are expressed, with *CSSACPD-C* transcripts being the most abundant; three of the seven *CSFAD2* genes are expressed, with *CSFAD2A* being the highest; all three of the *CSFAD3* genes are expressed, but of these, only *CSFAD3A* increases during embryo development, with *CSFAD3B and CSFAD3C* present at very low levels. *CSD8* and *CSD6* show similar low levels of expression up until the U stage with transcripts of both genes being absent at the later FND stage (Figure 2a). Based on homology and expression analysis, lead candidates for each of the desaturation steps shown in Figure 2b can be identified as *SACPD-C*, *CSFAD2A*, *CSFAD3A* and *CSD6*. We focused our efforts on functionally characterizing *CSFAD2A* and *CSFAD3A*.

Characterization of the *Cannabis sativa* microsomal desaturase *CSFAD2A*

Quantitative RT-PCR analysis confirmed high-level expression of *CSFAD2A* during embryo development, peaking at the FND stage where it was more than 1000 times higher than in young

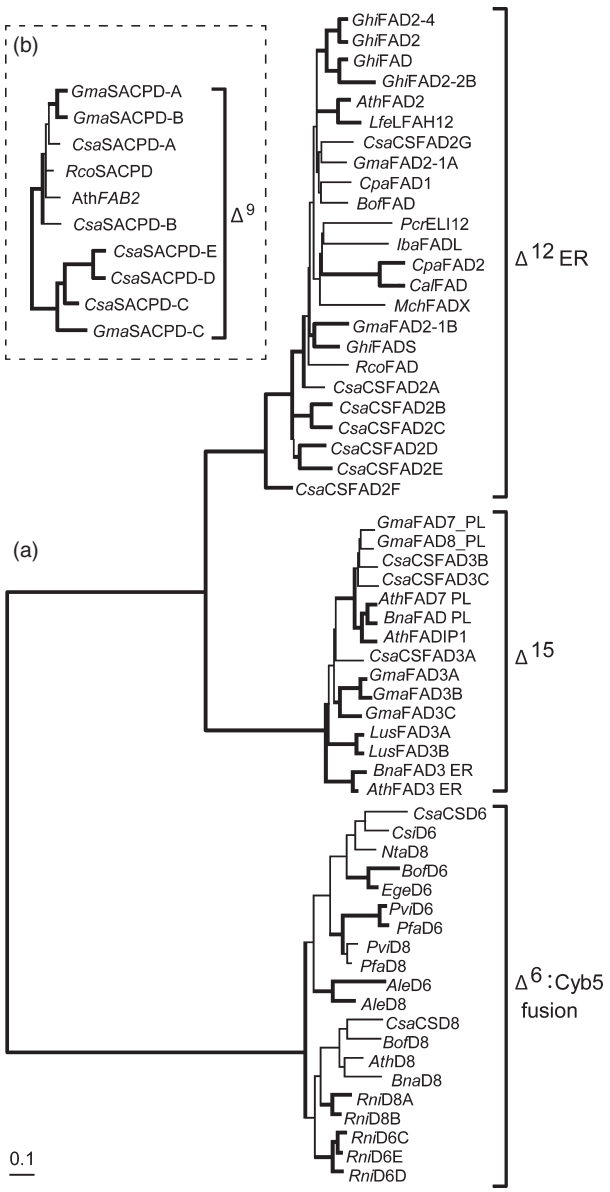

Figure 1 Evolutionary relationship of *Cannabis sativa* putative desaturases (retrieved from the Purple Kush genome) with known desaturases from selected species. (a) Genes encoding membrane-bound Δ6, Δ8 sphingo-lipid, Δ12 and Δ15 desaturases. (b) Soluble Δ9 Stearoyl-ACP desaturases. Species-specific identifiers: *Anemone lendsquerelli* (*Ale*), *Arabidopsis thaliana* (*Ath*), *Brassica napus* (*Bna), Borago officinalis* (*Bof*), *Crepis alpine* (*Cal*), *Crepis palaestina* (*Cpa*), *Cannabis sativa* (*Csa*), *Camellia sinensis* (*Csi*), *Echium gentianoides* (*Ege*), *Gossypium hirsutum* (*Ghi*), *Glycine max* (*Gma*), *Impatiens balsamina* (*Iba*), *Lesquerella fendleri* (*Lfe*), *Linum usitatissimum* (*Lus*), *Momordica charantia* (*Mch*), *Nicotiana tabacum* (*Nta*), *Petroselinum crispum* (*Pcr*), *Primula farinosa* (*Pfa*), *Primula vialli* (*Pvi*), *Ricinus communis* (*Rco*), *Ribes nigrum* (*Rni*). All branches are drawn to scale as indicated by the scale bar (0.1 substitutions/site). Strongly supported nodes with above 70% bootstrap values are highlighted with thickened lines.

leaves (Figure 3a). A similar pattern of expression but at much lower levels was observed for the *CSFAD2B* gene with the difference in expression between leaves and embryo much less pronounced, being about 20 times higher at the FND stage (Figure 3a). To confirm the functional identity of *CSFAD2A*, we cloned the corresponding ORF into the expression vector *pESC-TRP* containing the galactose-inducible GAL1 promoter and heterologously expressed this in *Saccharomyces cerevisiae*. This yeast has been used successfully for functional expression of several plant microsomal desaturases, because it acts as a convenient host with a simple fatty acid profile due to the presence of only a Δ9 desaturase producing palmitoleate and oleate, and the appropriate redox chain in a suitable membrane (Reed *et al.*, 2000). Fatty acid analysis of transformed yeast cells revealed the presence of two new fatty acids that were not present in either wild-type yeast or the empty vector control (Figure 3b, Table 1). Gas chromatography (GC) analysis of fatty acid methyl esters (FAMEs) demonstrated that the major novel peak is linoleic acid. As shown in Table 1, 72% of the endogenous oleic acid ($18{:}1^{\Delta 9}$) appears to have served as substrate for CSFAD2A and been converted into linoleic acid ($18{:}2^{\Delta 9,12}$), confirming CSFAD2A to have Δ12 desaturase activity. We transesterified the FAME fraction to 3-pyridylcarbinol esters and used GCMS to identify the second novel peak as ($16{:}2^{\Delta 9,12}$; Figure S5). We conclude that CSFAD2A can also use palmitoleic acid ($16{:}1^{\Delta 9}$) as substrate, with a conversion efficiency to $16{:}2^{\Delta 9,12}$ of 43% (Figure 3b, Table 1). Feeding eicosenoic acid ($20{:}1^{\Delta 11}$) to *CSFAD2A*-transformed yeast cultures resulted in 62% conversion to $20{:}2^{\Delta 11,14}$ (Figure S6) demonstrating that the enzyme can accept 16–20 C fatty acids and that the specificity is most accurately described as Δ*x*+3 (Schwartzbeck *et al.*, 2001).

Identification and characterization of two *CSFAD2A* desaturase mutants

To establish the *in vivo* role of *CSFAD2A*, we screened an ethyl methane sulphonate (EMS) mutagenized M2 outcrossed population of Finola using the TILLING method (Till *et al.*, 2006). We identified an allelic series of mutations among which *csfad2a-1* carries a stop codon at amino acid position 167. We performed two rounds of backcrossing of *csfad2a-1* to Finola and obtained homozygous *csfad2a-1* individuals (BC_2F_1) by crossing heterozygous male and female BC_2 siblings. *csfad2a-1* homozygotes displayed a dramatic increase in oleic acid content to 77 molar% in seed oil (Figure 3c, Table S2). In parallel, the levels of LA and ALA were strongly decreased compared to the fatty acid profile of the segregating wild-type seed oil from the same population, suggesting that this decrease was at the expense of the increase in oleic acid (Figure 3c, Table S2). Two novel fatty acids appeared in *csfad2a-1* at five and two molar per cent (Table S2). GC retention times indicated these to be 18:2 fatty acids and GCMS following derivatization to 3-pyridylcarbinol esters revealed these to be $18{:}2^{\Delta 6,9}$ and $18{:}2^{\Delta 9,15}$, respectively (Table S2). These may arise through the action of other desaturases on the high percentage oleic acid present in the developing embryos of *csfad2a-1*. The dramatic fatty acid level changes observed in *csfad2a-1* seed confirmed that the predicted truncated CSFAD2A protein is nonfunctional. Interestingly, no major changes in seed fatty acid profile were observed if the mutation was present in the heterozygous state, indicating that only one copy of this highly expressed *CSFAD2A* gene is sufficient to maintain the near wild-type level of fatty acids in hemp seed.

We also identified a second allele, *csfad2a-2*, which carries two point mutations giving rise to a proline to leucine transition at positions 218 and 375 of the predicted amino acid sequence

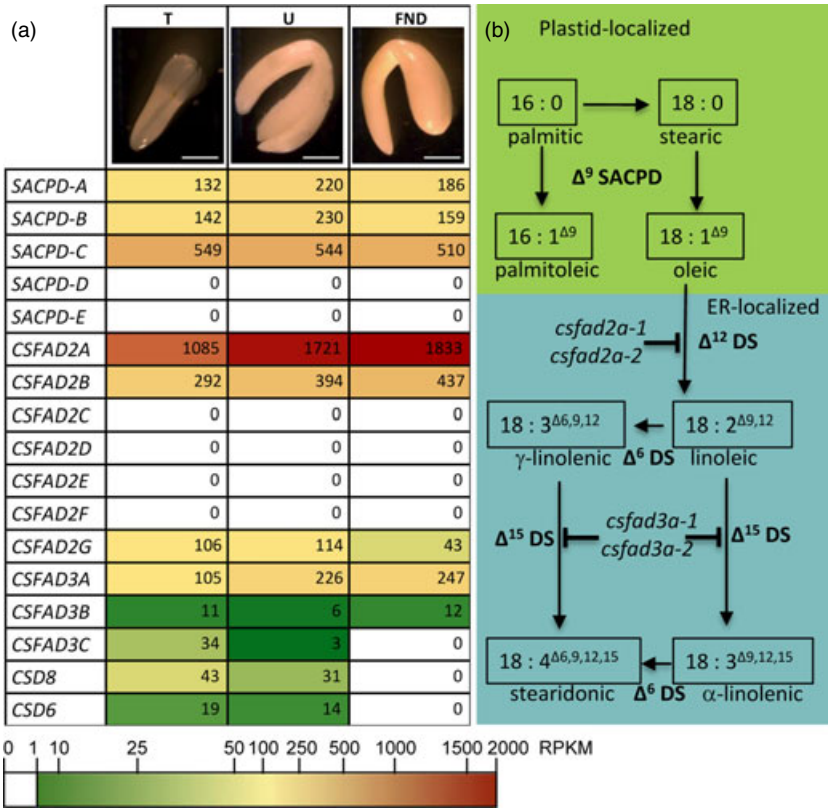

Figure 2 Expression of putative desaturase genes in developing embryos of the hemp cultivar Finola and metabolic context. (a) Embryos representative of each developmental stage used for RNA isolation are shown (Scale bar = 1 mm). Gene expression is depicted in a heat map format with RPKM values included. (b) Schematic presentation of the biosynthetic pathway giving rise to the major fatty acids in hemp seed oil. SACPD—stearoyl-ACP desaturase, DS—desaturase. Enzymatic steps are shown in bold and those steps compromised by mutation in specific *CSFAD2* and *CSFAD3* genes as detailed in Figures 3 and 4 are indicated.

of CSFAD2A. Homozygous *csfad2a-2* (BC_1F_1) seed accumulate nearly 70 molar per cent of oleic acid, low levels of 18:$2^{\Delta6,9}$ and 18:$2^{\Delta9,15}$ and decreased levels of LA and ALA compared to heterozygous and segregating wild-type seeds from the same population (Figure 3d, Table S2). This seed oil phenotype is very similar to that of *csfad2a-1* (Figure 3c) and is consistent with one or both of the P to L transitions disrupting protein function. This is expected given the importance of proline amino acids in determining protein structure. Interestingly, the levels of oleic acid, linoleic acid and α-linolenic acid remained unchanged in leaf tissues of both *csfad2a-1* and *csfad2a-2* compared to wild-type plants (Table S3), which is consistent with the gene expression data showing *CSFAD2A* to be largely seed specific (Figure 3a).

Characterization of the *Cannabis sativa* microsomal desaturase *CSFAD3A*

Quantitative RT-PCR confirmed expression of *CSFAD3*A in both leaves and embryos and showed it to be induced during seed development peaking at the FND stage where it is about 14 times higher than levels in young leaves (Figure 4a). Heterologous expression of *CSFAD3A* in *S. cerevisiae* followed by fatty acid feeding resulted in desaturation of linoleic (18:$2^{\Delta9,12}$) to α-linolenic acid (18:$3^{\Delta9,12,15}$) and γ-linolenic acid (18:$3^{\Delta6,9,12}$) to stearidonic acid (18:4$\Delta^{6,9,12,15}$) at a conversion efficiency of 56% and 23%, respectively (Figure 4b, Table 1). The yeast *CSFAD3A* transformants also exhibited low level activity with endogenous 16:$1^{\Delta9}$ and 18:$1^{\Delta9}$ resulting in what we identified as 16:$2^{\Delta9,15}$ and 18:$2^{\Delta9,15}$, respectively (Table 1, Figures S5 and S7). *CSFAD3A* transformants did not show any activity on exogenously supplied 20:$1^{\Delta11}$ after 28 h incubation. Together these results confirm that CSFAD3A acts as a Δ15 desaturase when expressed in *S. cerevisiae*.

Identification of mutations in *CSFAD3* confirms Δ15 desaturase activity

We screened our EMS-mutagenized hemp population and identified an allelic series of mutations in *CSFAD3* including one, designated *csfad3a-1*, that results in a stop codon being introduced at codon position 255. We performed three rounds of backcrossing to Finola and obtained homozygous *csfad3a-1* (BC_3F_1) seeds by crossing BC_3 siblings. Seed oil of the homozygous *csfad3a-1* contained near zero and zero levels of ALA and SDA, respectively, an elevation of LA from 55 to 75 molar per cent and no significant effect on GLA compared to the segregating wild type and heterozygotes in the M5 generation (Figure 4c, Table S2). A similar seed oil phenotype was seen in BC_2F_1 material (Table S2). These dramatic changes in the homozygous *csfad3a-1* seed oil profile confirmed that CSFAD3A acts as a Δ15 desaturase *in vivo* as well as in a heterologous host. Interestingly, when the mutation is in the heterozygous state, an intermediate phenotype is displayed in the seed oil with just half the levels of ALA and SDA compared to wild type. A second mutant, *csfad3a-2*, carried a point mutation resulting in conversion of proline to leucine at amino acid position 190 and this resulted in a similar seed oil phenotype to *csfad3a-1* (Figure 4d, Table S2). In contrast to seed oil, the production of ALA in the leaf tissue of both *csfad3a-1* and *csfad3a-2* is decreased by only 6% and 7%, respectively, compared to wild type (Table S3). This suggests the expression of other genes encoding Δ15 desaturase enzymes in leaf tissue, with *CSFAD3B* and *CSFAD3C* being obvious candidates.

High oleic hemp oil product performance

We selected *csfad2a-1* for further analysis, extended the backcrossing to generate BC_4 material and bulked up *csfad2a-1* seed

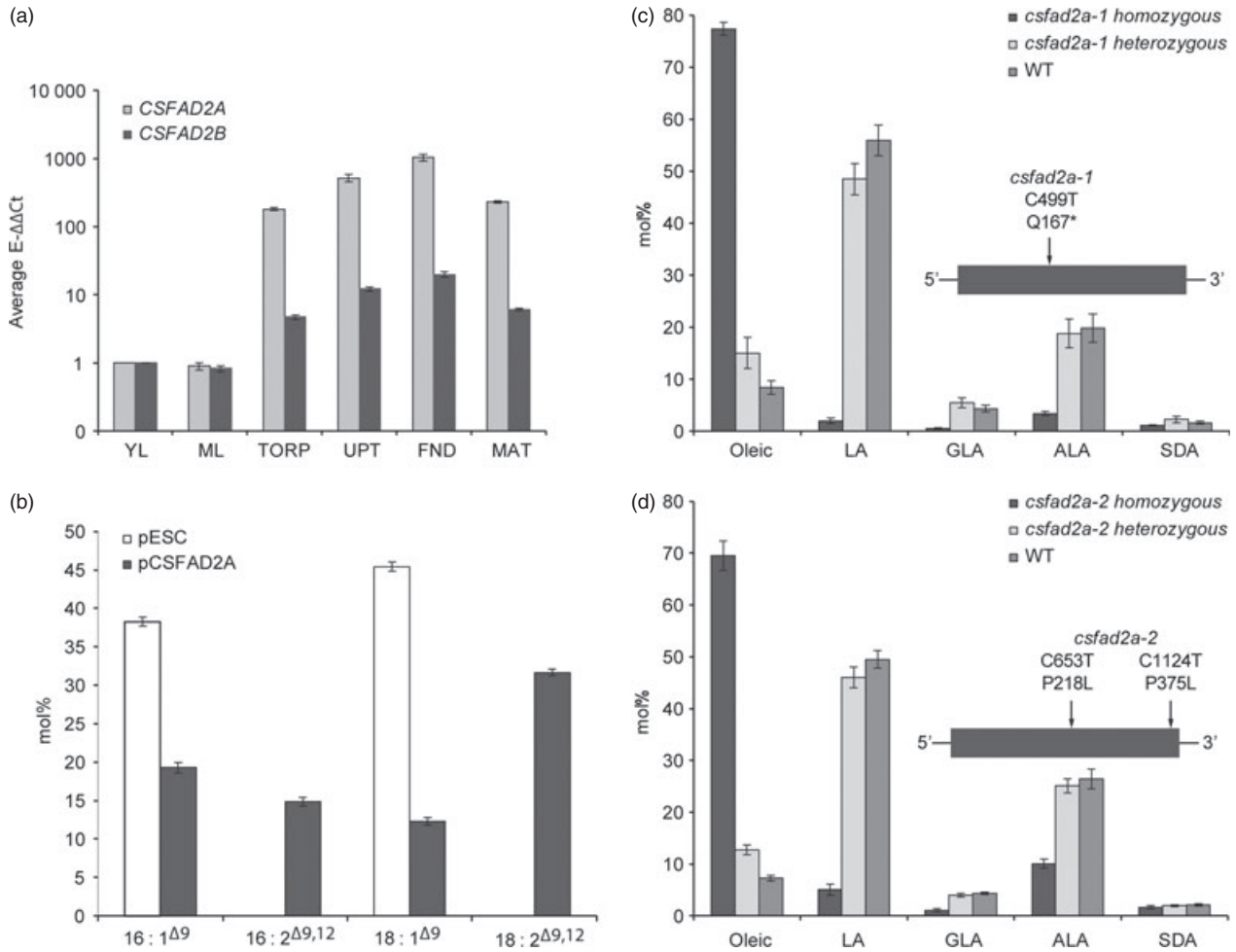

Figure 3 Characterization of *CSFAD2A* gene function. (a) Expression of *CSFAD2A* and *CSFAD2B* in developing embryo and mature leaf tissue compared to levels in young hemp leaves. Mean values represent the average of three biological replicas each consisting of three technical replicates. Abbreviations: YL: young leaves; ML: mature leaves; TORP: torpedo stage of embryo; UPT: upturned stage of embryo; FND: filled-not-desiccated stage of embryo; MAT: mature seed embryo. (b) Fatty acid composition of *Saccharomyces cerevisiae* transformed with either *CSFAD2A* cDNA or an empty vector (pESC-TRP) control. Each value is the mean ± SD from three independent experiments. Fatty acid composition of seed oil from (c) homozygous *csfad2a-1* (BC_2F_1) and (d) homozygous *csfad2a-2* (BC_1F_1) plants compared to respective segregating heterozygous and wild-type plants from the same generation as detailed in Table S2. Each value is the mean ± SD from 8 to 28 seeds from same line and generation.

by crossing homozygous mutant siblings. This material, now referred to as 'High Oleic Hemp' was grown in a single block field trial in Yorkshire, UK, during the 2011 growing season. Overall plant growth habit, flowering time and seed yield per plant were similar to the Finola wild type. Seed was cold-pressed giving a percentage oil of approximately 36% in the wild type and *csfad2a-1* material (Figure 5a). Fatty acid composition analysis confirmed the high oleic status of cold-pressed field grown *csfad2a-1* material, on a par with a commercial high oleic rapeseed material (Figure 5b). Rancimat determination of oxidation stability of the pressed oil is an industry standard methodology that allows shelf life to be determined by extrapolation of oxidation at elevated temperatures. We found that high oleic hemp *csfad2a-1* oil had an increased shelf life from 1.5 to 10 months at 20 °C, 4.1 to 28.6 months at 4 °C and 5.3 to 37.1 months at 0 °C (Figure 5c). Shelf life of high oleic rapeseed oil is also longer than standard rapeseed oil (Figure 5c), but shelf life of the high oleic hemp oil exceeds that calculated for high oleic rapeseed oil despite them having equivalent amounts of oleic and polyunsaturated fatty acids (Figure 5b, c). Plant seeds contain antioxidants such as tocopherols, which are thought to play a role in preventing oxidation of polyunsaturated fatty acids. We measured levels of tocopherols in high oleic hemp oil and found these to be significantly higher than that present in Finola hemp oil (Figure 5d), and also significantly higher than in both standard rapeseed oil and high oleic rapeseed oil (Figure S8). Consistent with the increased stability of high oleic hemp and rapeseed oils, we found that both produced decreased levels of volatile aldehydes as determined by head space analysis (Figure S9). Not surprisingly, the high oleic hemp TAG composition consisted mainly of triolein, which was completely absent from Finola hemp oil (Figure S10).

Discussion

Mutagenesis has been used to increase the amount of genetic variation available for selective breeding since the 1940s but as induced mutations are mostly recessive, resulting phenotypes are

Table 1 Fatty acid composition of yeast transformants and pESC-TRP empty vector controls fed with fatty acid substrates. Percentage conversion was calculated as product/(substrate + product) * 100 at the assay endpoint. Each value is the mean ± SD from three independent experiments

Substrate	Substrate endpoint mol% total fatty acids	Product	Product endpoint mol% total fatty acids	% conversion
pESC-TRP				
$16:1^{\Delta 9}$*	38.3 ± 0.6	–	–	–
$18:1^{\Delta 9}$*	45.4 ± 0.6	–	–	–
$18:2^{\Delta 9,12}$	10.9 ± 0.3	–	–	–
$18:3^{\Delta 6,9,12}$	14.1 ± 0.5	–	–	–
$20:1^{\Delta 11}$	0.8 ± 0.05	–	–	–
pCSFAD2A				
$16:1^{\Delta 9}$*	19.3 ± 0.7	$16:2^{\Delta 9,12}$	14.8 ± 0.6	43 ± 1.4
$18:1^{\Delta 9}$*	12.3 ± 0.5	$18:2^{\Delta 9,12}$	31.7 ± 0.4	72 ± 0.8
$20:1^{\Delta 11}$	0.3 ± 0.01	$20:1^{\Delta 11,14}$	0.5 ± 0.01	62 ± 0.3
pCSFAD3A				
$16:1^{\Delta 9}$*	38.2 ± 0.3	$16:2^{\Delta 9,15}$	1.6 ± 0.04	3.9 ± 0.1
$18:1^{\Delta 9}$*	38.0 ± 0.8	$18:2^{\Delta 9,15}$	3.6 ± 0.1	8.7 ± 0.4
$18:2^{\Delta 9,12}$	4.2 ± 0.2	$18:3^{\Delta 9,12,15}$	5.4 ± 0.2	56.3 ± 0.5
$18:3^{\Delta 6,9,12}$	10.7 ± 0.8	$18:4^{\Delta 6,9,12,15}$	3.2 ± 0.2	23.1 ± 0.4
$20:1^{\Delta 11}$	0.59 ± 0.6	–	–	–

*Endogenous substrate; no fatty acid added to medium.

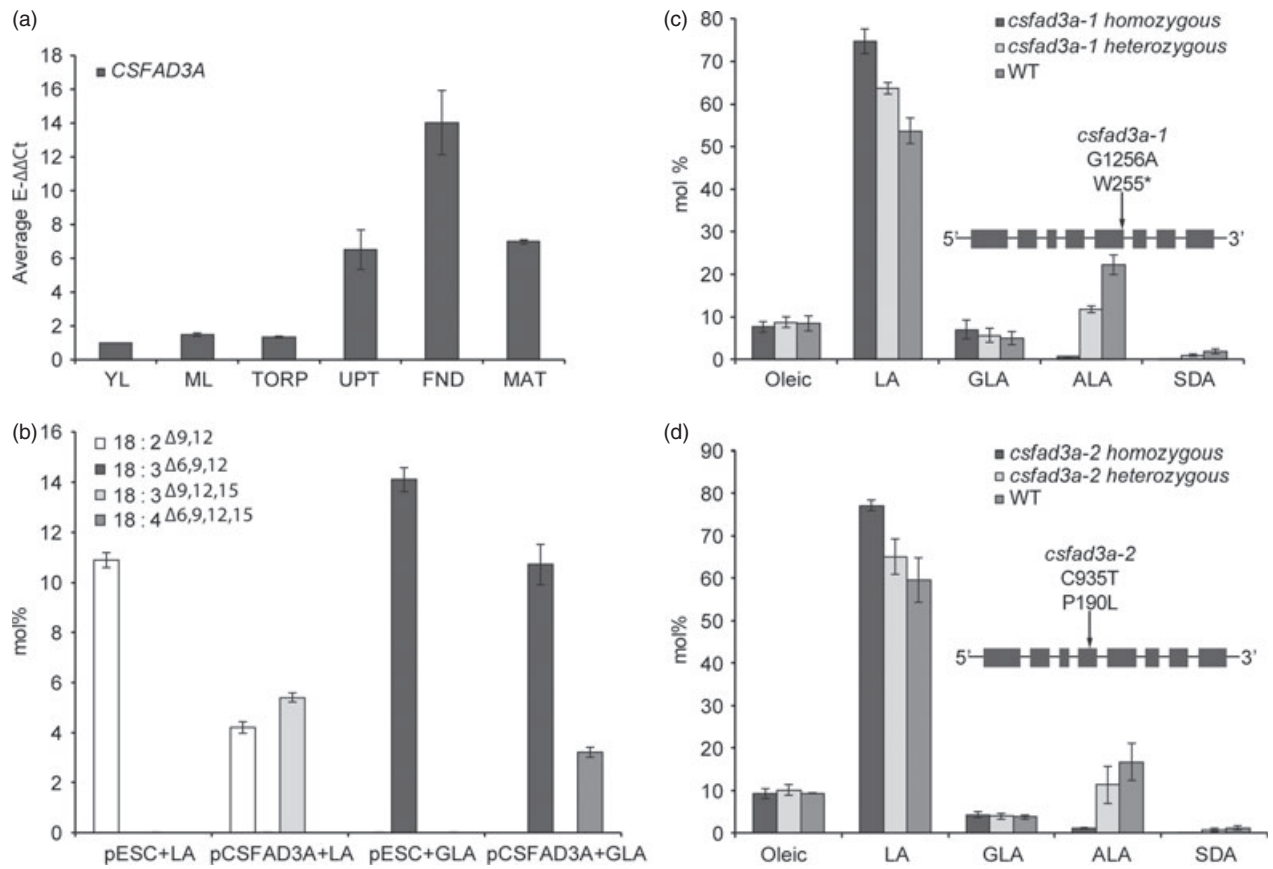

Figure 4 Characterization of *CSFAD3A* gene function. (a) Expression of *CSFAD3A* in developing embryo and mature leaf tissue compared to levels in young hemp leaves. Mean values represent the average of three biological replicates each consisting of three technical replicates. Abbreviations: YL: young leaves; ML: mature leaves; TORP: torpedo stage of embryo; UPT: upturned stage of embryo; FND: filled-not-desiccated stage of embryo; MAT: mature seed embryo. (b) Fatty acid composition of *Saccharomyces cerevisiae* transformed with either *CSFAD3A* cDNA or an empty vector (pESC-TRP) control. Each value is the mean ± SD from three independent experiments. Fatty acid composition of seed oil from (c) homozygous *csfad3a-1* (BC_3F_1) and (d) homozygous *csfad3a-2* (BC_2F_1) plants compared to respective segregating heterozygous and wild-type plants from the same generation as detailed in Table S2. Each value is the mean ± SD from 4 to 20 seeds from same line and generation.

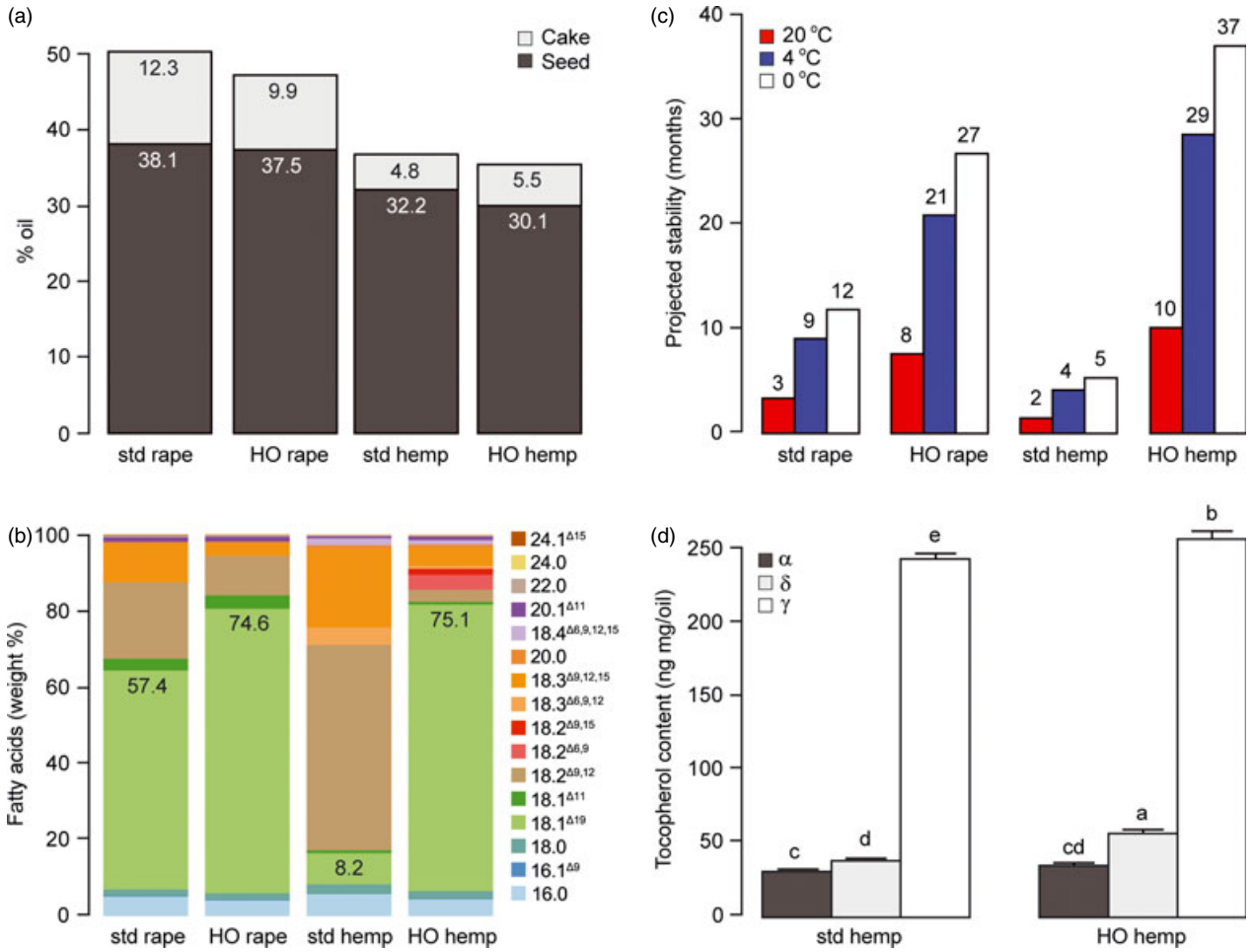

Figure 5 Cold-pressed oil analyses from standard (std) or high oleic (HO) hempseed and rapeseed. Small batches of seed harvested from field plots (~150 g) were cold-pressed and analysed for total oil content in the cake and seed (a), relative distribution of fatty acids (b) and rancimat-assayed stability at three different temperatures (c). Tocopherol assays are shown for hemp seed only (d). All data are representative assay values taken from the second or third pressed oil batches after the press had been preconditioned with appropriate seed and reached uniform operating temperatures. For tocopherol analyses (d), values are means ± 1 standard error from five analyses from the same oil batches with letters above bars indicating significantly different groups (ANOVA and Tukey's HSD; $P < 0.05$).

only observed when the mutant allele is in the homozygous state. In dioecious species such as hemp, obtaining homozygous mutations is a particular challenge, hence the TILLING (Till *et al.*, 2006) method, which allows identification of mutations in the heterozygous state, is particularly valuable. The advent of next generation sequencing technologies permits cost-effective, routine identification of genes from species not previously described at the molecular level and the present work demonstrates that comprehensive target gene selection can give rise to predictable breeding outcomes even in such species.

Heterologous expression in *S. cerevisiae* revealed that CSFAD2A desaturates $16{:}1^{\Delta 9}$ to $16{:}2^{\Delta 9,12}$ and $18{:}1^{\Delta 9}$ to $18{:}2^{\Delta 9,12}$ at 43% and 72% efficiency, respectively (Figure 3b, Table 1). However, the fact that the heterologously expressed protein also desaturates $20{:}1^{\Delta 11}$ eicosenoate to $20{:}2^{\Delta 11,14}$ demonstrates that this desaturase does not have strict Δ12 regiospecificity. Different integral membrane desaturases have evolved at least three distinct counting mechanisms for positioning double bonds as previously summarized (Shanklin and Cahoon, 1998). In addition to enzymes that count from either the carboxyl end or the methyl end of the molecule, there are also examples of enzymes that appear to follow the Δ*x*+3 rule whereby, rather than measuring from the carboxyl or methyl ends, the desaturase references the existing double bond at position '*x*' and places the new double bond (methylene interrupted) at position *x*+3 towards the methyl group (Hitz *et al.*, 1994; Schwartzbeck *et al.*, 2001). For example, the FAD2 enzyme from developing seeds of peanut (*Arachis hypogaea* L.) which, in addition to using palmitoleate and oleate as substrates also converts $19{:}1^{\Delta 10}$ to $19{:}2^{\Delta 10,13}$, leading the authors to conclude that it follows the Δ*x*+3 rule. This rule also best describes the consistent placement of the second double bond in the three monounsaturated substrates used by CSFAD2A.

Heterologous expression of CSFAD3A demonstrates that it introduces a double bond at the Δ15 position of $18{:}2^{\Delta 9,12}$ producing $18{:}3^{\Delta 9,12,15}$ and $18{:}3^{\Delta 6,9,12}$ producing $18{:}4^{\Delta 6,9,12,15}$ at conversion efficiencies of 56.3 and 23.1%, respectively (Table 1). This enzyme also produces $18{:}2^{\Delta 9,15}$ and $16{:}2^{\Delta 9,15}$ from $18{:}1^{\Delta 9}$ and $16{:}1^{\Delta 9}$ but at lower conversion efficiencies of 3.9 and 8.7%. Unlike CSFAD2A, CSFAD3A does not use $20{:}1^{\Delta 11}$ as substrate. Thus, while efficient introduction of a Δ15 double bond by CSFAD3A requires a substrate with Δ9 and Δ12 double bonds

consistent with previous reports (Shanklin and Cahoon, 1998), it can also use both 16C and 18C substrates carrying only a Δ9 double bond but at lower efficiency. We therefore conclude that CSFAD3 exhibits Δ15 regioselectivity. The ability to use oleic acid as substrate, albeit at lower conversion efficiency than linoleic acid or γ-linolenic acid, provides an explanation for the appearance of the $18{:}2^{\Delta 9,15}$ at two molar per cent in the high oleic seed oil of *csfad2a-1* (Table S2). These results are consistent with the observation that overexpression of a FAD3 in seeds of Arabidopsis leads to production of a polymethylene-interrupted dienoic fatty acid (Puttick *et al.*, 2009). A similar explanation involving an endogenous Δ6 activity could account for the presence of small amounts of $18{:}2^{\Delta 6,9}$ in *csfad2a* mutants.

The seed oil phenotypes that we observe for both *csfad2a* and *csfad3a* are largely consistent with the activity data obtained from heterologous expression. Production of high levels of oleic acid in both *csfad2a-1* and *csfad2a-2* confirm that this gene is responsible for the major Δ12 activity in developing hemp seeds as depicted in Figure 2b. Small amounts of LA, ALA, GLA and SDA totalling seven molar per cent in the *csfad2a-1* non-sense mutant (Figure 3c, Table S2) suggest a low level of Δ12 desaturase activity remains in developing seeds of this mutant. This residual activity could be encoded by *CSFAD2B,* which is also expressed in developing seeds but at much lower levels than *CSFAD2A* (Figure 3a). In contrast, leaf fatty acid composition is unaffected in *csfad2a-1* and *csfad2a-2* (Table S3), consistent with the seed specific expression of the gene (Figure 3a).

The decrease of ALA ($18{:}2^{\Delta 9,12,15}$) from approximately 20 molar per cent in wild type to <1% in both alleles of the *csfad3a* mutant together with the near elimination of SDA ($18{:}4^{\Delta 6,9,12,15}$; Figure 4c, d, Table S2) confirms that *CSFAD3A* encodes the major Δ15 activity in developing hemp seeds. Whether or not an endogenous Δ6 desaturase activity contributes to the relatively low levels of SDA in wild-type seeds as depicted in Figure 2b cannot be ascertained because the *csfad3a* mutant has severely decreased levels of the ALA ($18{:}3^{\Delta 9,12,15}$) substrate. The increase in LA ($18{:}2^{\Delta 9,12}$) from 55 to 75 molar per cent in seed oil of *csfad3a* mutants accounts for the decrease in ALA. However, this increase in LA has no effect on GLA ($18{:}2^{\Delta 6,9,12}$) levels, suggesting that the Δ6 desaturation of LA to GLA is limited by Δ6 enzyme rather than substrate availability. Consistent with this, the transcript levels of both the *CSD8* and *CSD6* are very low from the T to U stages of embryo development and are not detected at the FND stage (Figure 2a). Up-regulation of *CSD6* may therefore be the best strategy to increase amounts of GLA as has been demonstrated by genetic engineering of tobacco (Reddy and Thomas, 1996; Sayanova *et al.*, 1999a), *Brassica juncea* (Hong *et al.*, 2002), *Brassica napus* (Liu *et al.*, 2001) and evening primrose (de Gyves *et al.*, 2004).

Initial field trial results from backcrossed *csfad2a-1* material confirm high-level production of oleic acid in the field (Figure 5a). Rancimat assays showed an approximately sevenfold increase in high oleic hemp oil stability at three different temperatures compared to Finola oil (Figure 5c). High oleic rapeseed oil showed approximately twofold increase in stability compared to the control at three temperatures and overall, high oleic hemp was more stable despite the two oils having a similar percentage of oleic acid (Figure 5b). The major cause of increased oil stability is likely due to the decrease in relatively unstable polyunsaturated fatty acids and increase in the relatively stable oleic acid. In addition, it is tempting to speculate that the higher overall levels of tocopherols in high oleic hemp compared to high oleic rapeseed oil could account for the increase in stability of the former (Figure S8). Interestingly, there is a small but significant increase in levels of tocopherols in high oleic hemp oil compared to Finola oil (Figures 5d and S8), which could be due to either increased synthesis or decreased degradation of these antioxidants in high oleic oil seeds.

The availability of the complete suite of desaturase genes together with the ability to generate allelic series of mutations in candidate genes by TILLING makes Finola hemp an excellent choice for further investigation of seed oil metabolism as well as an important target for crop improvement.

Experimental procedures

cDNA library construction and EST preparation from developing seeds of *Cannabis sativa*

U stage tissue was ground to a fine powder in liquid nitrogen and RNA extracted using the RNAeasy kit (Qiagen, Hilden, Germany). cDNA was synthesized with the Creator™ SMART™ cDNA Library Construction Kit (Clontech, Mountain View, CA) and cloned into the pDNR-LIB vector (Clontech). 1893 ESTs were generated by Single-pass Sanger sequencing and a Blast similarity search identified two unigene sequences with homology to FAD2 desaturases. Random amplification of cDNA ends (RACE) was performed to obtain full-length sequences using primers for *CSFAD2A*: 5′-AAAATGGGAGCCGGTGGCCGAAT-3′ and 5′-GGGCGGAATTGCTTTCTTGATTTCGC-3′; RACE primers for *CSFAD2B*: 5′-GCAGACGATATGACCGTTTCGCTTCTCA-3′ and 5′-GCGAGTTGGTACAACACGAATGTGGTGA-3′.

To generate the *FAD3* gene from hemp cDNA, the following degenerate primers were used 5′-ACNCAYCAYCARAAYCAYGG-3′ and 5′-CAYTGYTTNCCNCKRTACCA-3′. To obtain full-length *CSFAD3*, the following RACE primers were used: 5′-CACGGCCATGTTGAGAATGACGAG-3′ and 5′-GGACAAACAGACAAGCAAAGCAGCCA-3′.

Genome mining, sequence alignment and phylogenetic analysis

Membrane-bound and soluble desaturase sequences were retrieved by local tBLASTn searches of the genome sequences from *C. sativa* varieties Purple Kush (canSat3) and Finola (Finola1), both of which were downloaded from the Cannabis Genome Browser (http://genome.ccbr.utoronto.ca/downloads.html; van Bakel *et al.*, 2011), using plant membrane-bound Δ6, Δ8 sphingo-lipid, Δ12 (FAD2) and Δ15 (FAD3), as well as the soluble Δ9 stearoyl-ACP desaturases as queries. Initial gene models of these desaturases were predicted with the FGENESH software (Salamov and Solovyev, 2000), readjusted by multiple sequence alignment using ClustalX (Thompson *et al.*, 1994) and further clarified by comparing their genomic coding sequences with corresponding EST sequences where possible. Undetermined residues were denoted as X for the corresponding N base calls in the nucleotide sequences based on the most homologous sequence in the multiple sequence alignment. Assembled nucleotide and predicted amino acid sequences for all identified desaturases are provided in Table S1 and are available from the Cannabis Genome Browser website: http://genome.ccbr.utoronto.ca/. The full-length and cDNA sequences of the functionally characterized CSFAD2A and CSFAD3A have been deposited in GenBank (Accession numbers KF679783 and KF679784, respectively).

All desaturase sequences from other plant species were retrieved from GenBank (http://www.ncbi.nlm.nih.gov/), these include Δ6 desaturases and Δ8 sphingo-lipid desaturases (Song *et al.*, 2010), Δ12 and Δ15 desaturases (Andreu *et al.*, 2010; Bilyeu *et al.*, 2003; Li *et al.*, 2008; Sperling and Heinz, 2003; Vrinten *et al.*, 2005; Zhang *et al.*, 2009), and the soluble Δ9 stearoyl-ACP desaturases (Zhang *et al.*, 2008). Sequences have been assigned a three-letter species-specific identifier as follows: *Anemone lendsquerelli* (Ale), *Arabidopsis thaliana* (Ath), *Brassica napus* (Bna), *Borago officinalis* (Bof), *Crepis alpine* (Cal), *Crepis palaestina* (Cpa), *Cannabis sativa* (Csa), *Camellia sinensis* (Csi), *Echium gentianoides* (Ege), *Gossypium hirsutum* (Ghi), *Glycine max* (Gma), *Impatiens balsamina* (Iba), *Lesquerella fendleri* (Lfe), *Linum usitatissimum* (Lus), *Momordica charantia* (Mch), *Nicotiana tabacum* (Nta), *Petroselinum crispum* (Pcr), *Primula farinosa* (Pfa), *Primula vialli* (Pvi), *Ricinus communis* (Rco), *Ribes nigrum* (Rni).

GenBank accession numbers for the protein sequences are as follows: Δ6, Δ8 sphingo-lipid desaturases: AleD6/8 (AAQ10731/AAQ10732), AthD8 (NP_191717), BnaD8 (CAA11857), BofD6/8 (AAC49700/AAG43277), CsiD6 (AAO13090), EgeD6 (AAL23580), NtaD8 (ABO31111), PfaD6/8 (AAP23034/AAP23033), PviD6/8 (AAP23036/AAP23035), RniD8A/B (ADA60228/ADA60229), and RniD6C/D/E (ADA60230/ADA60231/ADA60232); Δ12 desaturases: AthFAD2 (AAA32782), BofFAD (AAC31698), CalFAD (CAA76158), CpaFAD1 (CAA76157), CpaFAD2 (CAA76156), GhiFAD2-4 (AAQ16653), GhiFAD2 (AAL37484), GhiFAD (CAA71199), GhiFAD2-2B (ABY71269), GhiFADS (CAA65744), GmaFAD2-1A/B (AAB00860/ABF84062), IbaFADL (AAF05915), LfeLFAH12 (AAC32755), MchFADX (AAF05916), PcrELI12 (AAB80697), RcoFAD (AAC49010); Δ15 desaturases: AthFAD3_ER (P48623), AthFAD7_PL (BAA03106), AthFADIP1_PL (BAA04504), BnaFAD_PL (AAA61774), BnaFAD3_ER (P48624), GmaFAD3A/B/C (AAO24263/AAO24264/AAO24265), GmaFAD7/8 (ACF19424/ACU17817), LusFAD3A/3B (ABA02172/ABA02173); Δ9 stearoyl-ACP desaturases: AthFAB2 (AEC10310), GmaSACPD-A/B/C, (AAX86050/AAX86049/ABM45911), RcoSACPD (AAA74692).

All deduced amino acid sequences from *C. sativa* were aligned with desaturases of other plant species using ClustalX. The alignments were reconciled and further adjusted by eye to minimize insertion/deletion events, and only the most conserved alignment regions were used in the subsequent phylogenetic analyses. Distance analyses used the Protdist program with a Jones-Taylor-Thornton substitution matrix of the Phylip 3.6b package (Felsenstein, 1989). Phylogenetic trees were calculated from the distance matrices by the neighbour-joining algorithm. Bootstrap analyses consisted of 1000 replicates using the same protocol. Groups with above 70% bootstrap value were considered as strongly supported.

Deep sequencing the developing hemp seed transcriptome

Pyrosequencing was carried out on three cDNA libraries prepared from dissected embryos at the T, U and FND stages at the GenePool genomics facility at the University of Edinburgh on the 454 GS-FLX sequencing platform (Roche Diagnostics, Branford, CT). Raw sequence analysis, contiguous sequence assembly and annotation were performed as described (Graham *et al.*, 2010). Abundance of membrane-bound and soluble desaturase transcripts were analysed *in silico* by determining read counts in the three EST libraries. The raw reads were mapped to the reference sequence, which consisted of the open reading frames of the 17 desaturase genes (included in Table S1) with BWA mapping software (Li and Durbin, 2009). The raw read counts were retrieved from the resulting output file for each gene in the libraries, and the counts were then normalized to a reads per kilobase per million reads (RPKM) value as an approximation of gene expression.

Quantitative real-time PCR

Total RNA from leaves of 2-week-old and 4-week-old hemp plants was extracted with TRI Reagent Solution (Ambion®; Life Technologies, Carlsbad, CA) and treated with Turbo DNA-*free* (Ambion®). Single-strand cDNA was synthesized from DNase-treated RNA using SuperScript II (Invitrogen™; Life Technologies, Carlsbad, CA) reverse transcriptase with oligo(dT)$_{16–18}$ primer (Invitrogen) and diluted to 50 ng/μL. Quantitative real-time PCR was performed using an ABI Prism 7300 detection system (Applied Biosystems®; Life Technologies, Carlsbad, CA) and SYBR Green PCR Master mix (Applied Biosystems®) to monitor dsDNA synthesis. Gene-specific primers: 5′CTCGGACATAGGGATTTTCATTG3′ and 5′CAACCCAACCTAACCCTTTGG3′ for *CSFAD2A*; 5′CGATGTGGGGGTTTTCATCA-3′ and 5′-AACCCAACTCAACCCTCTTGCT-3′ for *CSFAD2B*; 5′TCAAATCCCACACTACCATCTTGT3′ and 5′TTTCTAGGCTCCCTGTAATACTTTCC3′ for *CSFAD3*. Amplification plots were analysed with an R_n threshold of 0.2 to obtain C_T (threshold cycle) values. Each transcript was normalized to the hemp *actin-2* gene amplified with primers: 5′GGGTCACACTGTGCCAATCTAC3′ and 5′CCCAGCAAGGTCAAGACGAA3′ and compared among samples.

Establishment and screening of an EMS-mutagenized population

Hemp seed of the Finola variety were purchased from the Finola company (http://www.Finola.com), Finland, and grown in controlled glasshouse facilities at the University of York. The seed was treated with 300 mM EMS for 5 h before sowing onto soil-based John Innes Compost No. 2. Mutagenized M1 female plants were outcrossed with male wild-type Finola plants to produce a heterozygous M2 screening population. Five nanogram per microlitre DNA from individual plants was pooled fourfold for screening. A 1140-bp fragment of *CSFAD2A* was amplified in a two-step PCR amplification. The first step was carried out with unlabelled primers (5′CCCATTGCTTTAAACGCTCTCTAATTCGCT3′ (left) and 5′CACCCCTAACCACATTAAGCCATACCCCAT3′ (right)) on 12.5 ng pooled gDNA in 10 μL volumes. Labelling of the amplified gene fragment with infrared dyes occurred during the second PCR step, where a mixture of labelled and unlabelled primers was used for further amplification and simultaneous labelling using appropriately diluted product from the first PCR step as template (left primer labelled with IRDye 700, right primer labelled with IRDye 800 (MWG, Ebersberg, Germany), left primer labelled:unlabelled ratio = 3:2; right primer labelled:unlabelled ratio = 4:1).

A 1500-bp fragment of *CSFAD3A* was also amplified in a two-step PCR reaction using nonlabelled gene-specific primers: 5′CGCCATTCCTAAGCATTGTT3′ (left) and 5′ATAGTGGTCCTGGCTGATGC3′ (right) in the first step. As for the Δ12 desaturase fragment, labelling with infrared dyes occurred during the second PCR step but using 5′M13-tailed primers: 5′TGTAAAACGACGGCCAGTGGGCTGCTCAAGGAACCATGTTCT3′ (left) and 5′AGGAAACAGCTATGACCATCCTTGGTAGCTTCCACAAGATGG3′ (right) mixed with M13 primers labelled with IRDye 700 and IRDye 800. The ratios of labelled to unlabelled primers were as for the *CSFAD2A* fragment. Heteroduplex formation, CEL I nuclease

digestion and analysis on the LI-COR 4300 DNA sequencer platform were carried out as described by Till *et al.* (2006).

Cloning and expression of *Cannabis sativa CSFAD2A* and *CSFAD3A* in *Saccharomyces cerevisiae*

The open reading frames of *CSFAD2A* and *CSFAD3A* were amplified by PCR using Phusion Hot Start DNA polymerase (Finnzymes) and the following pairs of specific primers: 5′ATAGGATCCAAA**ATG**GGAGCCGGT3′ (left) and 5′GCCTC GAGC**CTA**AAACTTGTTTTTGTACC3′ (right) for *CSFAD2A* and 5′GGGGAATTCATA**ATG**ACAGAATCACATGC3′ (left) and 5′TAG CGGCCGCATA**CTA**CATTTGCTTGGC3′ (right) for *CSFAD3*. Both open reading frames were cloned into the pESC-TRP (Stratagene) and transformed into *S. cerevisiae* strain G175 (Gietz and Woods, 2002). Cultures were grown at 28 °C in the presence of 2% (w/v) raffinose and 1% (w/v) Tergitol NP-40 (Sigma-Aldrich, St. Louis, MO). Expression of the transgene was induced when OD_{600} reached 0.2–0.3 by supplementing galactose to 2% (w/v) and appropriate fatty acids were added to a final concentration of 50 μM. Incubation was carried out at 25 °C for four generations (28 h).

Fatty acid and oil analysis

Fatty acid methyl esters (FAMEs) were prepared by direct transmethylation of single seeds or ~10 mg oil samples (Browse *et al.*, 1986). FAME content was determined by gas chromatography with flame ionization detection (GC-FID; GC Trace Ultra, Thermoquest Separation Products, Manchester, UK). A 1-μL aliquot of FAMEs in hexane was injected into a 3-mm internal diameter FocusLiner containing glass wool (SGE, Milton Keynes, UK) at 230 °C in programmed flow mode with H_2 as carrier gas. The H_2 flow program was as follows: initial hold 0.3 mL/min for 0.1 min, then ramped at 5 mL/min^2 to 0.5 mL/min for the remainder of the run. The split ratio was maintained at 1 : 250, and a gas saver slow of 20 mL/min was initiated at 1.5 min into the run. Separation was achieved using a narrow-bore cyanopropyl polysilphenylene-siloxane capillary column (SGE BPX70; 10 m length × 0.1 mm internal diameter × 0.2 μm film thickness). FAMEs were separated using the following temperature program: initial hold 150 °C 0.1 min, then ramped at 16 °C/min to 220 °C, followed by cool down to initial conditions at 120 °C/min. The FID was run at 300 °C with air, H_2 and make-up N_2 gases flowing at 350, 35 and 30 mL/min, respectively. The signal was collected and peaks detected and integrated using ChromQuest 4.2 software (Thermo Electron Corporation, Manchester, UK). FAMEs were identified and quantified relative to the Supelco 37 component FAME mix (Sigma-Aldrich, Gillingham, UK).

Extracts containing FAMEs that did not coelute with standards or whose identity was unclear were concentrated and further derivatized to their 3-pyridylcarbinol esters (Dubois *et al.*, 2006), chromatographed using a longer, thicker-film BPX70 column using He as carrier gas with an extended thermal gradient, and 70 eV electron impact mass spectra generated using a Leco Pegasus IV mass spectrometer running ChromaTof 4.5 software (Leco, Stockport, UK). Under these conditions, retention time order was preserved as per the GC-FID analyses. Mass spectra were interpreted to localize dienoic double bond positions as described by Christie *et al.* (1987).

Phenotyping for fatty acid content was carried out on single cotyledons dissected from 2-days-old seedlings germinated on moist filter paper. The surviving seedlings were transferred to soil, grown and genotyped, and then, selected individuals were used for subsequent crosses.

Oil pressing was carried out using a small capacity Komet screw press (Model CA 59 G; IBG Monforts, Mönchengladbach, Germany), with a 6-mm press nozzle die and a screw speed of 20 r.p.m. Running temperature was checked with a digital thermometer inserted into the restriction die, with screw-press barrel temperature not exceeding 60 °C. After each sample, all press devices were cleaned and dried.

The oxidative stability of the pressed oils was determined using a Metrohm Rancimat model 743, according to AOCS Official Method Cd 12b-92. Briefly, the induction times (n = 4) for portions of oil (3.0 g) were determined at 100, 110 and 120 °C and 20 L/h air throughput. Projected shelf life stability was calculated by extrapolation of the relationship between the measured induction time and the temperature (Metrohm Application Bulletin No. 141/3e).

Tocopherol analysis

Tocopherols were measured in 100 mg/mL dilutions of pressed oils in methyl tertiary butyl ether. Aliquots (10 μL) were separated by a nonaqueous reverse phase HPLC method used for neutral lipid separation (Burgal *et al.*, 2008) and quantified by their UV absorbance at 297 nm against calibration curves of authentic standards (Sigma-Aldrich, Gillingham, UK).

Acknowledgements

Financial support for this work came from the UK Technology Strategy Board, reference TP/3/BIO/6/1/17190, EU 7th Framework Programme award, project number 311849, 'MultiHemp' and to the Centre for Novel Agricultural Products from the Garfield Weston Foundation. We thank Martin Farrow, ADM Erith, UK, for providing the Hi-Oleic rapeseed, the University of York, Department of Biology horticulture staff for expert plant care and Judith Mitchell for administrative support.

References

Andreu, V., Lagunas, B., Collados, R., Picorel, R. and Alfonso, M. (2010) The GmFAD7 gene family from soybean: identification of novel genes and tissue-specific conformations of the FAD7 enzyme involved in desaturase activity. *J. Exp. Bot.* **61**, 3371–3384.

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

Baud, S., Boutin, J.P., Miquel, M., Lepiniec, L. and Rochat, C. (2002) An integrated overview of seed development in *Arabidopsis thaliana* ecotype WS. *Plant Physiol. Biochem.* **40**, 151–160.

Bilyeu, K.D., Palavalli, L., Sleper, D.A. and Beuselinck, P.R. (2003) Three microsomal omega-3-fatty acid desaturase genes contribute to soybean linolenic acid levels. *Crop Sci.* **43**, 1833–1838.

Browse, J., McCourt, P.J. and Somerville, C.R. (1986) Fatty acid composition of leaf lipids determined after combined digestion and fatty acid methyl ester formation from fresh tissue. *Anal. Biochem.* **152**, 141–145.

Burgal, J., Shockey, J., Lu, C., Dyer, J., Larson, T., Graham, I. and Browse, J. (2008) Metabolic engineering of hydroxy fatty acid production in plants: RcDGAT2 drives dramatic increases in ricinoleate levels in seed oil. *Plant Biotechnol. J.* **6**, 819–831.

Callaway, J. (2004) Hempseed as a nutritional resource: an overview. *Euphytica*, **140**, 65–72.

Callaway, J.C. and Laakkonen, T.T. (1996) Cultivation of *Cannabis* oil seed varieties in Finland. *J. Int. Hemp Assoc.* **3**, 32–34.

Christie, W.W., Brechany, E.Y. and Holman, R.T. (1987) Mass spectra of the picolinyl esters of isomeric mono- and dienoic fatty acids. *Lipids*, **22**, 224–228.

Dubois, N., Barthomeuf, C. and Bergé, J.-P. (2006) Convenient preparation of picolinyl derivatives from fatty acid esters. *Eur. J. Lipid Sci. Technol*. **108**, 28–32.

Felsenstein, J. (1989) PHYLIP – Phylogeny inference package (Version 3.2). *Cladistics*, **5**, 164–166.

Gietz, R.D. and Woods, R.A. (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. In *Methods in Enzymology*, Vol. 350 (Christine, G. and Gerald, R.F., eds), pp. 87–96. Salt Lake City: Academic Press.

Graham, I.A., Besser, K., Blumer, S., Branigan, C.A., Czechowski, T., Elias, L., Guterman, I., Harvey, D., Isaac, P.G., Khan, A.M., Larson, T.R., Li, Y., Pawson, T., Penfield, T., Rae, A.M., Rathbone, D.A., Reid, S., Ross, J., Smallwood, M.F., Segura, V., Townsend, T., Vyas, D., Winzer, T. and Bowles, D. (2010) The genetic map of *Artemisia annua* L. identifies loci affecting yield of the antimalarial drug artemisinin. *Science*, **327**, 328–331.

de Gyves, E.M., Sparks, C.A., Sayanova, O., Lazzeri, P., Napier, J.A. and Jones, H.D. (2004) Genetic manipulation of γ-linolenic acid (GLA) synthesis in a commercial variety of evening primrose (*Oenothera* sp.). *Plant Biotechnol. J.* **2**, 351–357.

Hitz, W.D., Carlson, T.J., Booth, J.R.J., Kinney, A.J., Stecca, K.L. and Yadav, N.S. (1994) Cloning of a higher-plant plastid ω-6 fatty acid desaturase cDNA and its expression in a cyanobacterium. *Plant Physiol*. **105**, 635–641.

Hong, H., Datla, N., Reed, D.W., Covello, P.S., MacKenzie, S.L. and Qiu, X. (2002) High-level production of γ-linolenic acid in *Brassica juncea* using a Δ6 desaturase from *Pythium irregulare*. *Plant Physiol*. **129**, 354–362.

Kinney, A.J. (1998) Production of specialised oils for industry. In *Plant Lipid Biosynthesis: Fundamentals and Agricultural Applications*(Harwood, J.L., ed), pp. 273–286. Cambridge, UK: Cambridge University Press.

Lee, M., Lenman, M., Banaś, A., Bafor, M., Singh, S., Schweizer, M., Nilsson, R., Liljenberg, C., Dahlqvist, A., Gummeson, P.-O., Sjödahl, S., Green, A. and Stymne, S. (1998) Identification of non-heme diiron proteins that catalyze triple bond and epoxy group formation. *Science*, **280**, 915–918.

Li, H.-L. (1974) An archaeological and historical account of cannabis in China. *Econ. Bot*. **28**, 437–448.

Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

Li, L., Wang, X., Gai, J. and Yu, D. (2008) Isolation and characterization of a seed-specific isoform of microsomal omega-6 fatty acid desaturase gene (FAD2-1B) from soybean. *DNA Seq*. **19**, 28–36.

Liu, J.-W., Huang, Y.-S., DeMichele, S., Bergana, M., Bobik, E., Hastilow, C., Chuang, L.-T., Mukerji, P. and Knutzon, D. (2001) Evaluation of the seed oils from a canola plant genetically transformed to produce high level of γ-linolenic acid. In *γ-Linolenic Acid: Recent Advances in Biotechnology and Clinical Applications*(Huang, Y.-S. and Ziboh, A., eds), pp. 61–71. Champaign, IL: AOCS Press.

Ohlrogge, J. and Browse, J. (1995) Lipid biosynthesis. *Plant Cell*, **7**, 957–970.

Puttick, D., Dauk, M., Lozinsky, S. and Smith, M.A. (2009) Overexpression of a FAD3 desaturase increases synthesis of a polymethylene-interrupted dienoic fatty acid in seeds of *Arabidopsis thaliana* L. *Lipids*, **44**, 753–757.

Reddy, A.S. and Thomas, T.L. (1996) Expression of a cyanobacterial Δ6-desaturase gene results in γ-linolenic acid production in transgenic plants. *Nat. Biotechnol*. **14**, 639–642.

Reed, D.W., Schäfer, U.A. and Covello, P.S. (2000) Characterization of the *Brassica napus* extraplastidial linoleate desaturase by expression in *Saccharomyces cerevisiae*. *Plant Physiol*. **122**, 715–720.

Salamov, A.A. and Solovyev, V.V. (2000) *Ab initio* gene finding in Drosophila genomic DNA. *Genome Res*. **10**, 516–522.

Sayanova, O., Davies, G.M., Smith, M.A., Griffiths, G., Stobart, A.K., Shewry, P.R. and Napier, J.A. (1999a) Accumulation of Δ6-unsaturated fatty acids in transgenic tobacco plants expressing a Δ6-desaturase from *Borago officinalis*. *J. Exp. Bot*. **50**, 1647–1652.

Sayanova, O., Shewry, P.R. and Napier, J.A. (1999b) Histidine-41 of the cytochrome *b*5 domain of the borage Δ6 fatty acid desaturase is essential for enzyme activity. *Plant Physiol*. **121**, 641–646.

Schwartzbeck, J.L., Jung, S., Abbott, A.G., Mosley, E., Lewis, S., Pries, G.L. and Powell, G.L. (2001) Endoplasmic oleoyl-PC desaturase references the second double bond. *Phytochemistry*, **57**, 643–652.

Shanklin, J. and Cahoon, E.B. (1998) Desaturation and related modifications of fatty acids. *Annu Rev Plant Physiol Plant Mol Biol*. **49**, 611–641.

Song, L.Y., Lu, W.X., Hu, J., Zhang, Y., Yin, W.B., Chen, Y.H., Hao, S.T., Wang, B.L., Wang, R.R. and Hu, Z.M. (2010) Identification and functional analysis of the genes encoding Δ6-desaturase from *Ribes nigrum*. *J. Exp. Bot*. **61**, 1827–1838.

Sperling, P. and Heinz, E. (2003) Plant sphingolipids: structural diversity, biosynthesis, first genes and functions. *Biochim. Biophys. Acta*, **1632**, 1–15.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. **22**, 4673–4680.

Till, B.J., Zerr, T., Comai, L. and Henikoff, S. (2006) A protocol for TILLING and Ecotilling in plants and animals. *Nat. Protoc*. **1**, 2465–2477.

Vrinten, P., Hu, Z., Munchinsky, M.-A., Rowland, G. and Qiu, X. (2005) Two FAD3 desaturase genes control the level of linolenic acid in flax seed. *Plant Physiol*. **139**, 79–87.

Zhang, P., Burton, J.W., Upchurch, R.G., Whittle, E., Shanklin, J. and Dewey, R.E. (2008) Mutations in a Δ9–Stearoyl-ACP-desaturase gene are associated with enhanced stearic acid levels in soybean seeds. *Crop Sci*. **48**, 2305–2313.

Zhang, D., Pirtle, I.L., Park, S.J., Nampaisansuk, M., Neogi, P., Wanjie, S.W., Pirtle, R.M. and Chapman, K.D. (2009) Identification and expression of a new delta-12 fatty acid desaturase (FAD2-4) gene in upland cotton and its functional expression in yeast and *Arabidopsis thaliana* plants. *Plant Physiol. Biochem*. **47**, 462–471.

Supporting information

Additional Supporting information may be found in the online version of this article:

Figure S1 Alignment of the predicted amino acid sequences of FAD2 proteins.
Figure S2 Alignment of the predicted amino acid sequences of plastidal and endoplasmic reticular FAD3 desaturases.
Figure S3 Alignment of the predicted amino acid sequences of plant Δ6 desaturases.
Figure S4 Alignment of the predicted amino acid sequences of plant stearoyl-acyl carrier protein desaturases.
Figure S5 Hexadecadienoic acid double bond localization.
Figure S6 Eicosadienoic acid double bond localization.
Figure S7 Octadecadienoic acid double bond localization.
Figure S8 Comparison of tocopherol content in standard and high oleic rapeseed and hemp seed oil.
Figure S9 Comparison of small molecule volatiles emitted from standard and high oleic rapeseed and hemp seed oil.
Figure S10 Triacylglycerol analysis reveals an abundance of triolein in High Oleic Hemp Oil.
Table S1 Nucleotide, amino acid sequences, exon number and source of soluble and membrane-bound desaturases from *Cannabis sativa*.
Table S2 Fatty acid content in seed oil of hemp *csfad2a* and *csfad3a* mutants.
Table S3 Fatty acid content in leaves of hemp *csfad2a* and *csfad3a* desaturase mutants.

