The Plant Journal (2006) **45**, 847–856 doi: 10.1111/j.1365-313X.2005.02636.x

TECHNICAL ADVANCE

A high-throughput screen for genes from castor that boost hydroxy fatty acid accumulation in seed oils of transgenic Arabidopsis

Chaofu Lu[†], Martin Fulda[‡], James G. Wallis and John Browse[*]
Institute of Biological Chemistry, Washington State University, Pullman, WA 99164-6340, USA

Received 16 August 2005; revised 13 October 2005; accepted 31 October 2005.
[*]For correspondence (fax +1 509 335 7643; e-mail jab@wsu.edu).
[†]Present address: Department of Plant Sciences and Plant Pathology, Montana State University, Bozeman, MT 59717, USA.
[‡]Present address: Department of Plant Biochemistry, Georg-August University, Goettingen, Justus-von-Liebig Weg 11, D37077 Goettingen, Germany.

Summary

It is desirable to produce high homogeneity of novel fatty acids in oilseeds through genetic engineering to meet the increasing demands of the oleo-chemical industry. However, expression of key enzymes for biosynthesis of industrial fatty acids usually results in low levels of desired fatty acids in transgenic oilseeds. The abundance of derivatized fatty acids in their natural species suggests that additional genes are needed for high production in transgenic plants. We used the model oilseed plant *Arabidopsis thaliana* expressing a castor fatty acid hydroxylase (FAH12) to identify genes that can boost hydroxy fatty acid accumulation in transgenic seeds. Here we describe a high-throughput approach that, in principle, can allow testing of the entire transcriptome of developing castor seed endosperm by shotgun transforming a full-length cDNA library into an FAH12-expressing Arabidopsis line. The resulting transgenic seeds were screened by high-throughput gas chromatography. We obtained several lines transformed with castor cDNAs that contained increased amounts of hydroxy fatty acids in transgenic Arabidopsis. These cDNAs were then isolated by PCR and retransformed into the FAH12-expressing line, thus confirming their beneficial contributions to hydroxy fatty acid accumulation in transgenic Arabidopsis seeds. Although we describe an approach that is targeted to oilseed engineering, the methods we developed can be applied in many areas of plant biotechnology and functional genomic research.

Keywords: high-throughput, full-length cDNA library, transgenic plants, ricinoleate, castor, Arabidopsis.

Introduction

For several years scientists have sought to identify enzymes responsible for synthesis of derivatized fatty acids and to transfer the cognate genes into agronomically amenable crops. Such derivatized fatty acids have a considerable range of applications, particularly in the production of surfactants, paints, plasticizers and specialty lubricants. The results of these experiments have typically been disappointing, in most cases producing plant lines with very low yields of the desired fatty acids (Broun and Somerville, 1997; Cahoon *et al.*, 1999; Lee *et al.*, 1998; Suh *et al.*, 2002). For example, ricinoleic acid (12-hydroxyoctadec-*cis*-9-enoic acid; 18:1-OH) biosynthesis in castor (*Ricinus communis*) is catalyzed by the oleate Δ12-hydroxylase (FAH12) (van de Loo *et al.*, 1995). Heterologous expression of FAH12 in the model oilseed plant Arabidopsis produced only up to 17% hydroxy fatty acids in seed oils, mainly in the form of ricinoleate, along with minor constituents including densipolic acid (18:2-OH), lesquerolic acid (20:1-OH) and auricolic acid (20:2-OH), which are products of the FAD3 Δ15-desaturase in combination with elongation by the FAE1 condensing enzyme acting on ricinoleate (Broun and Somerville, 1997; Smith *et al.*, 2003). This level is much lower than that found in castor oil, where ricinoleate constitutes approximately 90% of total fatty acids. Other efforts to produce unusual

fatty acids, such as acetylenic, monoenoic, eleostearic and parinaric acids, have likewise found that accumulation of unusual fatty acids in transgenic seeds is far below the proportion found in their respective natural sources (Cahoon *et al.*, 1999; Lee *et al.*, 1998; Suh *et al.*, 2002). These results suggest that expressing the single catalytic enzymes required for unusual fatty acid biosynthesis is insufficient to create transgenic plants producing large amounts of these fatty acids in seed storage oil. As the unusual fatty acids occur in abundance in their natural sources, we believe that additional necessary components for increased accumulation of unusual fatty acids in transgenic plants can be obtained from the source species.

The FAH12 enzyme belongs to a large family of fatty acid modification enzymes that are related to the Arabidopsis oleate Δ12-desaturase (FAD2) protein, which is responsible for the synthesis of polyunsaturated fatty acids (Okuley *et al.*, 1994). Ricinoleic acid is formed by hydroxylation of oleic acid (18:1) that is esterified to the *sn*-2 position of phosphatidylcholine (PC) in the endoplasmic reticulum (Bafor *et al.*, 1991). Studies have found that hydroxy fatty acids are exclusively present in seed storage lipids, mostly in the form of triacylglycerols (TAG), and are absent from membrane phospholipids (Millar *et al.*, 2000). Apparently, plants have efficient mechanisms to remove unusual fatty acids from sites of synthesis and transfer them to storage as TAG, therefore it has been suggested (Jaworski and Cahoon, 2003; Singh *et al.*, 2005) that these enzymes are obvious targets for metabolic engineering, including phospholipase A2 (Bafor *et al.*, 1991) and various acyl-CoA dependent or independent acyltransferases of the storage TAG synthesis pathway (Weselake, 2005).

In parallel with our investigations of genes known to be involved in lipid metabolism, we also considered approaches that would, in principle, allow testing of all the genes expressed in developing castor endosperm, without any *a priori* assumptions about which gene functions might be beneficial. These approaches are based on our hypothesis that some proteins in castor seeds, whether directly or indirectly involved in TAG synthesis, may have co-evolved with the hydroxylase to facilitate processing 18:1-OH and other intermediate metabolites. Here we describe the construction of a suitable cDNA library and development of transformation vectors that allow shotgun expression of castor endosperm proteins in Arabidopsis seeds expressing the castor FAH12. The resulting transgenic seeds were screened for changes in hydroxy fatty acid composition by a high-throughput gas-chromatography approach. In our first trial of this approach, we have obtained several lines that contained significantly higher amounts of hydroxy fatty acids, and confirmed the effects of the castor cDNAs isolated from these lines by retransforming into FAH12-expressing Arabidopsis plants.

Results

Overview and screening strategy of a high-throughput approach

To screen castor genes that may boost hydroxy fatty acid accumulation in transgenic Arabidopsis, we designed a high-throughput approach (Figure 1). This approach consists of five steps: (i) construction of a full-length cDNA library from developing castor seeds; (ii) mass sub-cloning of the cDNA library faithfully into a binary plant-expression vector so that the cDNAs are under control of a seed-specific promoter; (iii) shotgun transformation of the cDNA library into an FAH12-expressing Arabidopsis line and screening

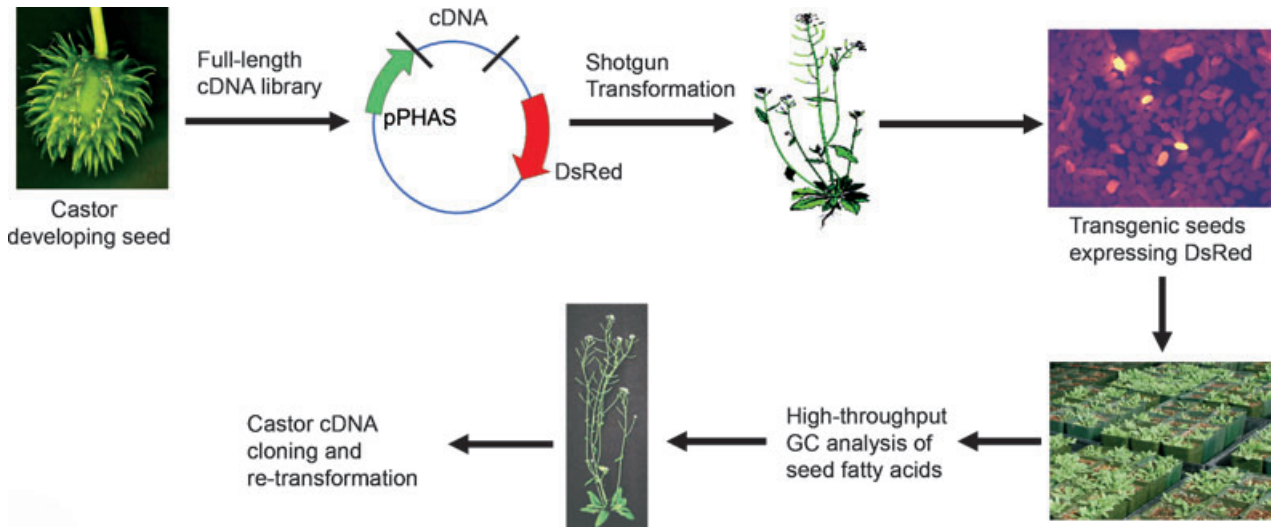

Figure 1. Outline of the high-throughput screen used to identify castor cDNAs that increase hydroxy fatty acid accumulation in transgenic seeds of Arabidopsis.

transgenic seeds using a fluorescent DsRed marker; (iv) analyzing fatty acid composition of T_2 transgenic seeds employing high-throughput gas chromatography (GC) to identify transgenic lines with increased hydroxy fatty acid content; and (v) isolation of castor cDNAs from these lines and confirmation of their beneficial effects by retransformation into the FAH-expressing Arabidopsis plants. The development of these steps is described in detail in the following sections.

Synthesis and selection of full-length cDNAs from developing castor endosperm

In order to efficiently test the function of individual genes from developing castor seeds directly in transgenic plants, it is highly desirable to use a cDNA library that is enriched in full-length clones. Reliance on a standard cDNA library, where perhaps 30% of the inserts are full length, would mean that fewer than 1/3 of the plants screened would contain a functional transgene. Castor bean stores oils in the seed endosperm. We therefore used developing endosperm as the starting material, and purified messenger RNA from endosperm 20 days after pollination at developmental stage IV, when the endosperm undergoes rapid dimensional growth and gain in weight (Greenwood and Bewley, 1982).

A critical step in making a full-length cDNA library lies in the synthesis of full-length first-strand cDNAs, but our first reverse transcription experiments produced first-strand cDNAs dominated by short, approximately 1.0-kb, partial-length cDNA fragments (data not shown). Several protocols that reportedly benefit long cDNA synthesis (Carninci *et al.*, 1998; Spiess and Ivell, 2002), including using trehalose or other ingredients in reverse transcription reactions, produced little improvement. However, we successfully improved reverse transcription by conducting the reaction at a high temperature (55°C) using the thermal-stable SuperScript III reverse transcriptase (Invitrogen), and the resulting single-strand cDNAs displayed a wide range of size distribution with the majority around 2–3 kb (Figure 2). This result suggested that the first-strand cDNAs we obtained most closely reflected the mRNA population in developing endosperm of castor.

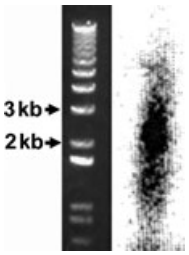

Figure 2. Full-length first-strand cDNA synthesis by reverse transcriptase reaction.
Lane 1, molecular weight markers; lane 2, [α-^{32}P]dATP radiolabeled cDNA.

We adopted a biotinylated cap-trapping approach to recover single-strand full-length cDNA molecules (Carninci *et al.*, 2000). This method is based on chemical introduction of a biotin group into the diol residue of the cap structure of mRNA, followed by RNase I treatment to destroy single-strand RNAs. The full-length cDNA–mRNA hybrids were selected by trapping the biotin residue at the cap sites using streptavidin-coated magnetic beads. Incompletely synthesized cDNAs were thus eliminated, because the hybrid molecules did not contain the cap structure after RNase I treatment destroyed the single-stranded mRNA. After purification of the first-strand cDNAs, the second-strand cDNA was synthesized by a highly efficient single-strand linker-ligation method, which uses DNA ligase to add a dsDNA linker to the 5′ ends of single-strand full-length cDNAs (Shibata *et al.*, 2001). This method offers great advantages over the traditional oligo(dG) tailing method (Carninci and Hayashizaki, 1999), as elimination of the GC tail simplifies sequencing and translation of proteins. We used linker sequences that contain the *Sst*I restriction site. An *Xho*I site had been incorporated into the primer sequence for first-strand cDNA synthesis. Therefore both ends of cDNA can be cleaved, allowing oriented cloning of double-strand cDNAs. The methylation-sensitive enzymes *Sst*I and *Xho*I do not cut hemi-methylated cDNAs, and we had used 5-methyl-dCTP instead of dCTP for the first-strand cDNA synthesis.

Vectors designed for shotgun transformation

We chose to design a bacteriophage lambda vector to minimize size bias and improve the likelihood of isolating rare clones (Short *et al.*, 1988). To facilitate subsequent mass sub-cloning for expression in plants, we created a vector designated λ_{GW}, incorporating the *attB* sites of the Gateway cloning system (http://www.Invitrogen.com/gateway) using the proven λ_{ZAPII} vector (Stratagene) (Short *et al.*, 1988) as the starting material (Figure 3a). The Gateway system allows directional sub-cloning of genes from one vector to any other compatible vectors by site-specific DNA recombination. Its ability to transfer DNA with high efficiency and fidelity makes it a suitable system for mass sub-cloning of an entire cDNA library. In addition, inclusion of a *ccdB* cell-death gene in a Gateway vector avoids high background in a cDNA library with clones having no inserts. Before investing effort in library construction, we tested our vector construct using a 2.0-kb cDNA clone, and found that the efficiency of our λ_{GW} vector was comparable with that of the λ_{ZAPII} vector, producing approximately 10^8 pfu μg^{-1} DNA.

We also created pGate-DsRed-Phas, a high-throughput binary vector for plant transformation (Figure 3b). This vector contains the *attR* sites so that cDNAs can be cloned

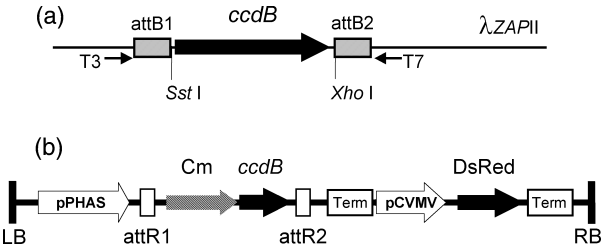

Figure 3. Simplified maps of library and transformation vectors.
(a) The λ_{GW} vector for cDNA library construction, which is derived from the λ_{ZAPII} vector by incorporating the Gateway system for mass sub-cloning; and *Sst*I and *Xho*I restriction sites for directional cloning.
(b) The T-DNA region of the binary plant expression vector pGate-DsRed-Phas.
attB and attR, Site-specific recombination sequences; ccdB, a topoisomerase poison from *Escherichia coli*; Cm, chloramphenicol resistance; LB, left border; pCVMV, CVMV promoter; pPHAS, phaseolin promoter; RB, right border; T3 and T7, PCR and sequencing primers; Term, phaseolin and CVMV terminators.

directionally under control of a seed-specific phaseolin promoter (Sengupta-Gopalan *et al.*, 1985) via the Gateway cloning process. The vector employs a DsRed marker under control of the cassava vein mosaic virus (CVMV) promoter, and permits ready identification of transgenic seeds by examination under green illumination using a red filter (Stuitje *et al.*, 2003). To test the efficiency of the expression vector, a full-length cDNA clone of the *FAH12* gene was inserted into the vector by the Gateway process, and was subsequently transformed into Arabidopsis plants mediated by *Agrobacterium tumefaciens*. Resulting transgenic seeds were readily identified on the basis of DsRed expression (Figure 1). Analysis of these DsRed-positive seeds confirmed that they produced up to 15% hydroxy fatty acids.

Evaluation of a full-length castor cDNA library

We cloned the full-length double-stranded cDNAs into the λ_{GW} vector, and obtained an unamplified primary cDNA library containing approximately 5×10^5 clones. To evaluate the quality of this cDNA library, we performed plaque PCR and sequenced the cDNA inserts. As shown in Figure 4, the insert sizes of the cDNAs range from approximately 700 bp to >6 kb, and the majority of the clones have inserts of 2–3 kb. This indicates that the clone population of the library reflects the first-strand cDNA population. We also randomly selected 140 clones and sequenced the PCR products from the 5′ ends. Inspection of sequences using the ExPASy translation tool (http://us.expasy.org/tools/dna.html) identified putative ATG start codons with upstream in-frame stops. The sequences were used to search in Arabidopsis and National Center for Biotechnology Information (NCBI) non-redundant databases using the BLASTX program. Of the 140 sequences, 132 were identified as full-length clones when compared with the 5′ ends of their putative full-length homologs in the databases. As an oligo(dT) primer corresponding to the poly(A) signal sequence of the mRNA was used to synthesize the first-strand cDNA, the 3′ ends of the cDNA clones should include sequences for carboxyl termini of proteins. We therefore estimate that the cDNA library has approximately 94% clones encoding full-length proteins from developing castor endosperm.

Two of the 140 clones were found to encode the *FAH12* gene. This high frequency indicates that the castor endosperm used for cDNA library construction was at the right stage for hydroxy fatty acid synthesis. Furthermore, we also determined by PCR analysis that many important genes involved in lipid biosynthesis are present in this cDNA library, such as the genes for long-chain acyl-CoA synthetases (LACS); lysophosphatidic acyltransferases (LPATs); acyl-CoA:diacylglycerol acyltransferases (DGAT1 and DGAT2); and a phospholipid:diacylglycerol acyltransferase (PDAT) (data not shown). We considered normalizing our library to reduce the frequency of cDNAs corresponding to highly expressed genes such as those encoding seed-storage proteins. However, normalization protocols can lead to loss of low-abundance clones, so we decided to use the original library for the experiments described here.

The cDNA library in the λ_{GW} vector was excised into the plasmid format and mobilized into the pGate-DsRed-Phas vector via the Gateway cloning process, to create a plant expression library. The expression library was then electroporated into *A. tumefaciens* for transforming Arabidopsis plants. To test for any biases during these steps, we randomly selected 30 Agrobacterium colonies and performed colony PCR. The results indicated that cDNA size distribution was similar to that of the original library (data not shown).

Figure 4. Evaluation of a castor full-length cDNA library.
The size distribution of cDNA inserts of randomly picked clones from the cDNA library was examined by plaque PCR using T3 and T7 primers.
Two lanes of molecular weight markers are included.

Shotgun cDNA library transformation and high-throughput screen

To simplify screening of transgenic seeds containing increased proportions of hydroxy fatty acids, we produced several hydroxylase-expressing transgenic plant lines in the Arabidopsis *fae1* mutant background. This mutant lacks the FAE1 condensing enzyme required for the synthesis of lesquerolic and auricolic acids. When transformed with FAH12, *fae1* mutants contain only two hydroxy fatty acids (ricinoleic acid, 18:1-OH and densipolic acid, 18:2-OH), and this simplifies analysis of seed fatty acid composition. To identify homozygous lines of FAH12-transgenics with consistent, high hydroxy fatty acid content, we screened over 50 transgenic plants. Seeds from hemizygous plants contain up to 15% hydroxy fatty acids (ricinoleic plus densipolic acids). We chose six lines for further analysis, and the homozygous seeds contained 17–19% hydroxy fatty acids. One transgenic line, CL37, consistently contained 17% hydroxy fatty acids, and germinated normally. This line was used as recipient for shotgun transformation of the castor cDNA library.

In the first trial of our screening approach, we generated about 5000 T_1 transgenic seeds in the CL37 background screened by DsRed expression. A total of 3840 T_1 plants germinated and grew to maturity. The T_2 seeds were harvested from each of these T_1 plants, and random samples of about 30 seeds from each line were aliquoted into 96-well metal plates. Seeds were crushed and their fatty acyl methyl esters were derived by trimethylsulfonium hydroxide for single-pass gas-chromatography analysis in a high-throughput manner (Figure 5a). Data analyses indicated that most lines contained 16–18% total hydroxy fatty acids, similarly to the CL37 parental line. However, we found that some lines contained higher or lower amounts of total hydroxy fatty acids (Figure 5b). We chose to characterize 18 lines containing apparent increases in hydroxy fatty acids. To verify the high-throughput results of these lines, we sorted the T_2 seeds into red seeds transformed with castor cDNA, and brown untransformed segregants. Fatty acyl methyl esters were derived from these seed samples and analyzed by GC to compare fatty acid compositions between red and brown seeds from each line. We confirmed that eight lines contained 19–22% total hydroxy fatty acids in red seeds –

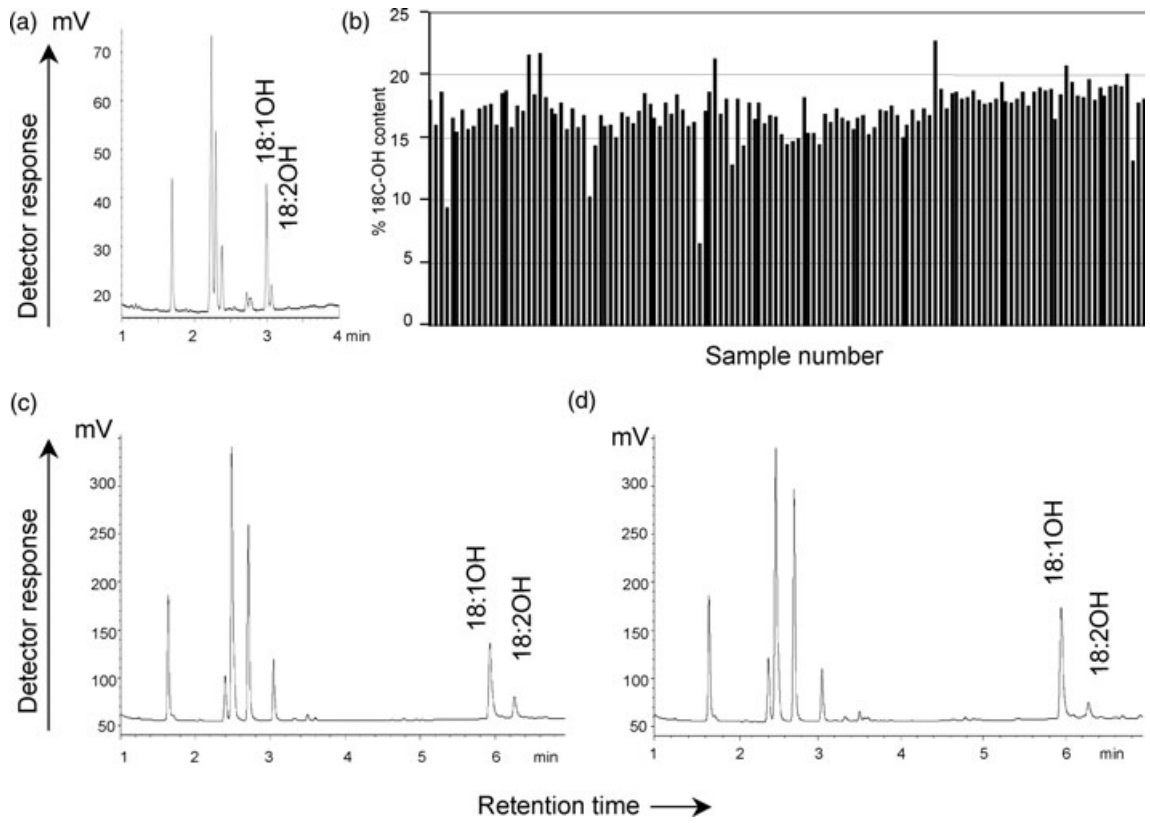

Figure 5. Examples of data analysis from the high-throughput screen for increased hydroxy fatty acid content.
(a, b) Seed fatty acid methyl esters from T_2 seeds of plants transformed with clones from a castor cDNA library were analyzed by gas chromatography (GC), the results were exported into Microsoft EXCEL, and the hydroxy fatty acid (18C-OH = 18:1OH + 18:2OH) contents were calculated and plotted for comparison.
(c) GC profile of fatty acid methyl esters of untransformed segregants from T_2 seeds.
(d) GC profile of the red-transformed T_2 segregant from the same T_2 seeds as in (b) with a castor oleosin cDNA containing increased hydroxy fatty acid content.

Table 1 Comparison of fatty acid composition between T_2 seeds transgenic with castor cDNA (upper rows) and untransformed segregants (lower rows)

Plant line	16:0[a]	18:0	18:1	18:2	18:3	18:1OH	18:2OH	Total hydroxy fatty acid	Increase (%)[b]
1204	Scaffold-attachment region DNA-binding protein								
	12.0 ± 0.2	6.0 ± 0.1	33.5 ± 0.3	20.6 ± 0.2	6.2 ± 0.1	17.2 ± 0.2	4.0 ± 0.04	21.2 ± 0.2	18.6
	15.4 ± 2.1	8.3 ± 0.7	31.2 ± 2.1	19.2 ± 1.0	5.4 ± 0.6	15.5 ± 0.3	2.4 ± 0.28	17.9 ± 0.3	
1938	Oleosin								
	12.6 ± 0.1	8.0 ± 0.1	30.9 ± 0.7	20.3 ± 0.6	6.4 ± 0.4	18.1 ± 0.4	3.3 ± 0.1	21.4 ± 0.3	19.7
	13.7 ± 0.8	8.2 ± 0.3	31.9 ± 2.6	21.0 ± 1.6	6.8 ± 0.2	15.3 ± 0.8	2.6 ± 0.5	17.9 ± 0.9	
1929	Phosphatidylethanolamine-binding protein								
	13.3 ± 0.4	6.3 ± 0.5	32.1 ± 0.8	22.4 ± 0.9	6.9 ± 0.4	15.4 ± 0.2	3.7 ± 0.2	19.1 ± 0.4	11.6
	14.5 ± 0.5	5.9 ± 0.2	31.9 ± 0.4	23.9 ± 0.1	6.7 ± 0.5	13.6 ± 0.2	3.5 ± 0.2	17.1 ± 0.4	

[a]Fatty acid data are mean ± SE of three separate sample preparations.
[b]Student's *t*-test significant at $P < 0.05$.

substantially more than that in the segregating untransformed seeds, which averaged 17% (Figure 5c,d).

Identification of castor genes and confirmation of their efficacy

The castor cDNA clones in eight lines that contained confirmed increases in hydroxy fatty acids were isolated by PCR, with genomic DNA from the transformed lines as template, and directed by primers corresponding to sequences flanking the castor cDNA in the pGate-DsRed-Phas vector. Single DNA fragments were amplified from three lines, indicating that only one cDNA had been transformed into each of these lines. Of the remaining five lines, four lines gave rise to two PCR products, and one to three PCR products, indicating that they contained multiple cDNA transgenes. Confirming the basis of the increased hydroxy fatty acid phenotype in these lines will require identification and testing of each of the cDNA clones separately, and perhaps in combination. This work is currently under way in our laboratory.

The cDNA sequences from the three single-insert lines were identified by database searches as homologs to oleosin (line 1938); phosphatidylethanolamine (PE)-binding protein (line 1929); and scaffold-attachment region (SAR) DNA-binding protein (line 1204) (Table 1). To confirm the effects of these castor genes on hydroxy fatty acid accumulation, we verified the complete sequence of the PCR products and cloned them into the pGate-DsRed-Phas vector via the Gateway system, then transformed them afresh into CL37 plants. The resulting DsRed-transformed seeds had similar increases in hydroxy fatty acids, confirming the positive effect of the genes on hydroxy fatty acid accumulation in transgenic Arabidopsis seed (data not shown).

Discussion

Plants produce many products with industrial applications, including oils, fibers, fuels and pharmaceuticals. The development of genetic engineering in plants offers new possibilities for improving the production of natural compounds, or for the synthesis of new products in plants. Over the past few years, it has been realized that the manipulation of single genes is of only limited value in metabolic engineering. As a result, strategies have been focused on more complex approaches involving simultaneous overexpression or suppression of multiple genes to achieve optimal metabolic flux (Capell and Christou, 2004). Understanding a metabolic network would facilitate the production of natural products and the synthesis of novel molecules in a predictable and useful manner.

Arabidopsis is a well studied model plant, and has been used for molecular dissection of many metabolic pathways, including those involved in lipid metabolism (Browse and Somerville, 1991). Heterologous expression of biosynthetic pathways in Arabidopsis is a powerful approach for developing metabolic engineering applications in plants. As an Arabidopsis plant will produce several thousand seeds under optimal glasshouse conditions, and as transformants can be obtained at a rate of 0.5–3% of all progeny seeds using a simple floral dip method, it is possible to obtain a large number of transgenic seeds in a few weeks. Here we describe a high-throughput, shotgun-transformation approach to identify castor genes that boost hydroxy fatty acid accumulation in transgenic Arabidopsis seeds. As outlined in Figure 1, to screen transformants in a high-throughput manner, several steps require high efficiency.

A full-length cDNA library is highly desirable for an expeditious functional test. In preparing a full-length cDNA library, we performed the reverse transcription reaction at a high temperature (55°C) instead of the traditional 37–42°C. This is expected to disrupt secondary structures that may be present in mRNAs, and thus greatly improve first-strand cDNA synthesis (Figure 2). We selected full-length cDNAs by an advanced cap-trapping method (Carninci and Hayashizaki, 1999) with great success: about 94% of the cDNA clones

in our cDNA library encode full-length proteins from developing castor endosperm, a frequency similar to that achieved by Carninci *et al.* (1996), who developed this cap-trapping method. By designing and constructing our own lambda vector, we anticipated that we would retain larger cDNA molecules in the library, and that these would be propagated successfully in *Escherichia coli* (Short *et al.*, 1988), reducing bias against large cDNA clones. Our results confirmed that the cDNA insert-size distribution in our library ranged from 700 bp to >6 kb, with majority of clones of 2–3 kb (Figure 4). This size distribution reflected that of the first-strand cDNAs (Figure 2).

To facilitate subsequent excision and mass sub-cloning into a binary plant expression vector, we incorporated the Gateway cloning system into the designs of our phage and plant transformation vectors. This technique not only avoided disruption of cDNA clones by restriction digestion, but also allowed cDNAs from the original lambda library to be transferred faithfully *en masse* into the transformation vector, and inserted directionally under control of the seed-specific phaseolin promoter (Figure 3b). In addition, the DsRed marker designed into our vectors allowed us to sort transformed from untransformed seeds easily. This ability greatly reduces the effort required to grow thousands of plants for high-throughput analysis, as every seed planted is known to be transgenic. The ability to differentiate transformed from untransformed seeds also speeds analysis by allowing separation of T_2 segregating seeds into those containing new castor transgenes from otherwise isogenic, untransformed segregants. Our expeditious high-throughput gas-chromatography method enabled us to analyze seed fatty acids composition in only 4 min per sample (see Experimental procedures). This protocol, when coupled with parallel injection using a dual-flame ionization detector (FID) GC, means that we could routinely analyze more than 600 seed samples per day.

The accumulation of hydroxy fatty acids depends on many factors, including the performance of the hydroxylase transgene and efficient channeling of hydroxy fatty acids into storage triacylglycerols. Our approach offers the opportunity to screen thoroughly all castor genes expressed in the developing endosperm. Our initial test screening of about 4000 transgenic plants resulted in eight lines that showed increased hydroxy fatty acid content. We identified three genes that occurred in the singly transformed lines, and confirmed their effects on hydroxy fatty acid accumulation. Oleosins are small proteins required for forming lipid bodies that contain storage triacylglycerols. We identified a castor oleosin as a genetic determinant mediating increased hydroxy fatty acid accumulation in transgenic Arabidopsis. This result suggests that overexpression of castor oleosins in the endoplasmic reticulum may facilitate the hydroxy fatty acid-containing TAGs being rapidly formed into oil bodies, thus preventing these unusual fatty acids from being incorporated into membrane lipids through the common diacylglycerol precursor. Phosphatidylethanolamine is one of the major phospholipids synthesized in the pathway shared by phospholipid and storage lipid biosynthesis. Although PC is generally regarded as the lipid substrate for fatty acid modification, fatty acids esterized to PE may also, to a lesser extent, be modified and channeled into TAG (Lin *et al.*, 2000). Recent reports indicate that including DNA sequences that encode SARs in transformation constructs may prevent potential silencing effects of transgenes (Allen *et al.*, 2000; Petersen *et al.*, 2002). It is possible that the SAR DNA-binding protein in line 1204 may have its effect by enhancing expression of the hydroxylase transgene, but other explanations are also possible. The precise roles of these genes in increasing hydroxy fatty acid accumulation will require further investigation.

The occurrence of very high ricinoleic acid content in castor seeds is probably the result of collective effects of several genes that have co-evolved to facilitate the incorporation of hydroxy fatty acids into TAGs. Although our initial screen has identified lines with moderate (approximately 20%) increases in hydroxy fatty acids, the results validate our hypothesis that this experimental approach can bring to our attention genes with hitherto unexpected roles in ricinoleate accumulation. The fact that several lines from the screen contain multiple inserts of castor cDNAs also suggests the possibility that our approach can identify combinations of cDNAs that increase hydroxy fatty acid accumulation. However, to improve the efficiency of the screen, one may wish to use a normalized or subtracted cDNA library, as some genes are very abundant in developing castor seeds (e.g. approximately 25% of the genes in our library encode storage proteins; C.L. and J.B., unpublished results). Also, to minimize homologous suppression of transgene expression, it may be advantageous to use different promoters to drive hydroxylase and castor cDNAs.

In conclusion, we describe a high-throughput approach capable of efficient discovery of genes for high-level accumulation of ricinoleate in transgenic plants. We created a transgenic line expressing the gene of interest in a model plant species, then shotgun transformed with a full-length cDNA library (which, in principle, may represent the complete transcriptome of the tissue where the gene is natively expressed), and examined their effects on accumulation of the desired gene product by a high-throughput gas chromatographic analytical method. Application of these principles to other genes of interest could help develop novel biotechnology-based crops with added value in nutrition, quality and yield. Although we describe an approach targeted at engineering oilseeds, the methods we have developed provide a powerful research tool in plant biotechnology and functional genomic research.

Experimental procedures

Construction of lambda and plasmid vectors

A bacterial phage vector λ_{GW} was designed to use the Gateway-mediated excision and transfer of cDNA inserts. The λ_{ZapII} vector (Stratagene, La Jolla, CA, USA) was modified by introducing the negative selection ccdB cassette using pDEST-C (Invitrogen, Carlsbad, CA, USA) as the amplification template. The primer pairs were: ccdBfor1, 5′-G*ACAAGTTGTACAAAAAAGCAGGCT* GAGCTCAGTATGCGTATTTGCGCGCTG-3′; ccdBfor2, 5′-AGCTG*ACAAGTTTGTACAAAAAAGCAGGCT*GAGCTC-3′; ccdBrev1, 5′-C*ACCACTTTGTACAAGAAAGCTGGGT*CTCGAGTACGCTAGTGTCATAGTCCTG-3′; ccdBrev2, TCGAC*ACCACTTTGTACAAGAAAGCTGGGT*CTCGAG. (Gateway sequences are shown in italics, *Sst*I and *Xho*I restriction sites are underlined.) The sticky-end PCR (Zeng, 1998) strategy used for this amplification introduced directional *attB* sites to achieve compatibility to the Gateway system, as well as the *Sst*I and *Xho*I restriction sites used later to clone the cDNA. The amplified cassette was ligated into λ_{ZapII} vector DNA arms prepared by published methods (Carninci and Hayashizaki, 1999). The ligation product was packaged with Max Plax (Epicentre, Madison, WI, USA) according to manufacturer's protocol. Cells of *Escherichia coli* strain C600 (F^- *thi*-1 *thr*-1 *leu*B6 *lac*Y1 *ton*A21 *sup*E44 λ^-), permissive for ccdB, were infected with the resulting phage particles, and single plaques were tested by PCR for presence of the ccdB cassette. A single positive plaque was selected and its sequence confirmed after amplification of the insert using standard M13 forward and reverse primers.

The expression vector pGate-DsRed-Phas was modified based on the pOEA2, a plant-transformation binary vector developed by Thomas Girke (Dow AgroSciences, San Diego, CA, USA). A fragment containing chloramphenicol resistance gene and ccdB cell-suicide gene bracketed by the Gateway *attR* sites was inserted between the seed-specific phaseolin promoter and terminator. A DsRed marker gene (Clontech, Mountain View, CA, USA) was placed behind the constitutive cassava vein mosaic virus (CVMV) promoter (Verdaguer *et al.*, 1996).

Plant material and transformation

We used an Arabidopsis mutant line, *fae1*, to generate transgenic plants expressing castor hydroxylase gene *FAH12*. One transgenic line, CL37, which stably expressed *FAH12* and contained approximately 17% of hydroxy fatty acids, was used to test cDNA clones from castor. These plants were transformed by the floral dipping procedure (Clough and Bent, 1998) following electroporation of individual cDNA clones or cDNA library into *Agrobacterium tumefaciens* strain GV3101 (pMP90). All plants were grown in controlled-environment chambers at 22°C under a 16-h photoperiod of 150 μmol quanta m^{-2} sec^{-1} photosynthetically active radiation.

Construction of a castor full-length cDNA library

Total RNA was extracted from developing castor endosperms using the Midi RNAqueous Kit (Ambion, Austin, TX, USA) according to the manufacturer's instructions. mRNA was prepared using the Oligotex kit (Qiagen, Valencia, CA, USA). mRNA (10 μg) was used for first-strand cDNA synthesis. The oligo(dT)BX primer containing *Bam*HI and *Xho*I restriction sites (underlined) was used: 5′-GAGAGAGA GAGAGAGA GAGGATCCACTCGAGTTTTTTTTTTTTTTTTVN-3′. The reverse-transcription reaction was performed as instructed by Invitrogen, using SuperScript III reverse transcriptase in a thin-well PCR tube by incubating at 55°C for 1 h in a thermocycler. We used [α-^{32}P]-dATP to monitor cDNA size distribution and calculate cDNA yield. Full-length first-strand cDNAs were precipitated and captured using the cap-trapping procedure as described by Carninci and Hayashizaki (1999). The second-strand cDNA was synthesized using the single-strand linker ligation method (Shibata *et al.*, 2001). The GN5 linker was prepared by mixing oligonucleotides A (5′-GAGAGA-GAGAGCACGAGCTCGTCGACTAGTGACACTATAGAACCAGNNNNN-3′) and B (5′-TGGTTCTATAGTGTCACTAGTCGACGAGCTCGTGC-TCTCTCTCTC-3′). The N6 linker was prepared by mixing oligonucleotides B and C (5′-GAGAGAGAGAGCACGAGCTCGTCGACT-AGTGACACTATAGAACCANNNNNN-3′). The reaction was followed by proteinase K treatment in the presence of 0.2% SDS and 10 mM EDTA at 45°C for 15 min. We added 1 μg glycogen, and extracted the reaction using phenol/chloroform and chloroform; the sample was precipitated using isopropanol. The double-stranded cDNA was digested with *Sst*I and *Xho*I and ligated into the λ_{GW} vector. The ligation product was packaged as described above.

For production of a plant expression library, the primary lambda library was amplified and DNA was extracted using a lambda DNA preparation and purification system (Promega, Madison, WI, USA). Up to 100 ng lambda DNA was mixed with 300 ng pDONR201 vector, 4 μl Gateway BP reaction buffer and 4 μl BP Clonase enzyme mix in a volume of 20 μl. Overnight incubation at 25°C was followed by proteinase K treatment and isopropanol precipitation. The precipitate was mixed with 300 ng *BstZ*17I-digested pGate-DsRed-Phas linearized vector, 4 μl Gateway LR buffer, and 4 μl LR Clonase enzyme mix in a volume of 20 μl. After overnight incubation at 25°C, the sample was further purified with proteinase K/phenol chloroform extraction, followed by ethanol precipitation. The resulting plasmids were electroporated into *Agrobacterium* strain GV3101. Colonies grown on Luria-Bertani (LB) plates supplemented with 100 mg l^{-1} spectinomycin (Sigma, St Louis, MO, USA) and 50 mg L^{-1} gentamicin (Fluka, St Louis, MO, USA) after 3 days' incubation at 30°C were scraped in a small volume of liquid LB medium. An aliquot was inoculated into 500 ml LB medium with the same antibiotics to prepare for transformation of Arabidopsis plants. The remaining cells were saved as glycerol stocks at −80°C.

Gas chromatography

For high-throughput GC analyses, about 30 seeds harvested from each mature T_2 transgenic line were aliquoted to custom 96-well stainless steel plates, and were cracked using a custom 96-tip handheld ball-crushing device. Fatty acyl methyl esters (FAMEs) were derivatized for 15 min in 100 μl trimethylsulfonium hydroxide solution (Muller *et al.*, 1993) and diluted with 100 μl methanol; both were delivered into the 96-well plates using a LEAP Twin-PAL dual-rail robot (LEAP Technologies, Carrboro, NC, USA) fitted to an Agilent 6890 dual-FID GC (Palo Alto, CA, USA). Then 0.3 μl sample was injected into the GC employing two 7 m × 0.53-mm Supelcowax columns (SPB50, Supelco, Bellefonte, PA, USA) using helium as the carrier gas. The GC analytical method was 190°C for 50 sec, 30°C min^{-1} ramp to 280°C, and the final temperature was maintained for 50 sec for a total run time of 4 min. Data were analyzed by graphical representation of hydroxy fatty acid composition in Agilent's CHEM STATION software and Microsoft EXCEL.

For confirmation of the high-throughput GC results and analysis of retransformed T_2 seeds, FAMEs were prepared by heating approximately 20 seeds at 80°C in 1 ml 2.5% H_2SO_4 (v/v) in methanol for 90 min, and extracted with 200 μl hexane and 1.5 ml 0.5% NaCl (w/v); 100 μl of organic phase were then transferred to auto-injector vials. Sample (1 μl) was injected into an Agilent 6890 GC fitted with a 15 m × 0.25-mm AT wax column (Alltech, Deerfield, IL, USA). The GC was programmed for an initial temperature of 190°C for 2 min,

followed by an increase of 8°C min^{-1} to 230°C, and maintained for a further 6 min.

Isolation of castor cDNA from transgenic plants and retransformation

DNA was extracted from young leaves by a hexadecyltrimethylammonium bromide (CTAB)-based method (Lukowitz *et al.*, 2000). Castor cDNAs in transgenic plants were amplified using primers CDIN1: 5′-ATCAACAAGTTTGTACAAAAAAGCAGGCTGAG-3′ and CDIN2: 5′-TCAACCACTTTGTACAAGAAAGCTGGGTCT-3′, and a hot-start DNA polymerase KOD (Novagen, Madison, WI, USA). The PCR conditions were 94°C for 3 min followed by 30 cycles of 94°C for 30 sec, 56°C for 30 sec and 72°C for 3 min, and finally 72°C for 10 min. The primer CDIN1 was also used for sequencing the 5′ ends of the amplified cDNA products using the BigDye sequencing kit (ABI, Foster City, CA, USA). For retransformation, cDNAs were cloned into the pGate-DsRed-Phas vector by the Gateway process and electroporated into *Agrobacterium* GV3101 strain for dipping flowers of Arabidopsis CL37 line, as described above.

Acknowledgements

We thank Julia Gothard, Bethany George-Carradine and Emily Fetcho for growing and harvesting plants; Dr Thomas Girke (Dow AgroSciences, San Diego) for constructing the pOEA2 vector used as the basis of our transformation vector; and Dr Rajesh Kumar for technical assistance with the high-throughput GC analysis. This work was supported through the Oilseed Engineering Alliance funded by Dow Chemical Company/DowAgrosciences LLC; by the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service grant no. 2003-35318-13914, and by the Agricultural Research Center at Washington State University.

References

Allen, G.C., Spiker, S. and Thompson, W.F. (2000) Use of matrix attachment regions (MARs) to minimize transgene silencing. *Plant Mol. Biol.* **43**, 361–376.

Bafor, M., Smith, M.A., Jonsson, L., Stobart, K. and Stymne, S. (1991) Ricinoleic acid biosynthesis and triacylglycerol assembly in microsomal preparations from developing castor bean (*Ricinus communis*) endosperm. *Biochem. J.* **280**, 507–514.

Broun, P. and Somerville, C. (1997) Accumulation of ricinoleic, lesquerolic, and densipolic acids in seeds of transgenic *Arabidopsis* plants that express a fatty acyl hydroxylase cDNA from castor bean. *Plant Physiol.* **113**, 933–942.

Browse, J. and Somerville, C. (1991) Glycerolipid synthesis: biochemistry and regulation. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **42**, 467–506.

Cahoon, E.B., Carlson, T.J., Ripp, K.G., Schweiger, B.J., Cook, G.A., Hall, S.E. and Kinney, A.J. (1999) Biosynthetic origin of conjugated double bonds: production of fatty acid components of high-value drying oils in transgenic soybean embryos. *Proc. Natl Acad. Sci. USA*, **96**, 12935–12940.

Capell, T. and Christou, P. (2004) Progress in plant metabolic engineering. *Curr. Opin. Biotechnol.* **15**, 148–154.

Carninci, P. and Hayashizaki, Y. (1999) High efficiency full-length cDNA cloning. *Methods Enzymol.* **303**, 19–44.

Carninci, P., Kvam, C., Kitamura, A. *et al.* (1996) High-efficiency full-length cDNA cloning by biotinylated CAP trapper. *Genomics*, **37**, 327–336.

Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. (1998) Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. *Proc. Natl Acad. Sci. USA*, **95**, 520–524.

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. (2000) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* **10**, 1617–1630.

Clough, S.J. and Bent, A.F. (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743.

Greenwood, J.S. and Bewley, J.D. (1982) Seed development in *Ricinus communis* (castor bean). I. Descriptive morphology. *Can. J. Bot.* **60**, 1751–1760.

Jaworski, J. and Cahoon, E.B. (2003) Industrial oils from transgenic plants. *Curr. Opin. Plant Biol.* **6**, 178–184.

Lee, M., Lenman, M., Banas, A. *et al.* (1998) Identification of non-heme di-iron proteins that catalyze triple bond and epoxy group formation. *Science*, **280**, 915–918.

Lin, J.T., Lew, K.M., Chen, J.M., Iwasaki, Y. and McKeon, T.A. (2000) Metabolism of 1-acyl-2-oleoyl-sn-glycero-3-phosphoethanolamine in castor oil biosynthesis. *Lipids*, **35**, 481–486.

van de Loo, F.J., Broun, P., Turner, S. and Somerville, C. (1995) An oleate 12-hydroxylase from *Ricinus communis* L. is a fatty acyl desaturase homolog. *Proc. Natl Acad. Sci. USA*, **92**, 6743–6747.

Lukowitz, W., Gillmor, C.S. and Scheible, W.R. (2000) Positional cloning in Arabidopsis. Why it feels good to have a genome initiative working for you. *Plant Physiol.* **123**, 795–805.

Millar, A.A., Smith, M.A. and Kunst, L. (2000) All fatty acids are not equal: discrimination in plant membrane lipids. *Trends Plant Sci.* **5**, 95–101.

Muller, K.D., Nalik, H.P., Schmid, E.N., Husmann, H. and Schomburg, G. (1993) Fast identification of *Mycobacterium* species by GC analysis with trimethylsulfonium hydroxide (TMSH) for transesterification. *J. High Resolut. Chromatogr.* **16**, 161–165.

Okuley, J., Lightner, J., Feldmann, K., Yadav, N., Lark, E. and Browse, J. (1994) *Arabidopsis Fad2* gene encodes the enzyme that is essential for polyunsaturated lipid synthesis. *Plant Cell*, **6**, 147–158.

Petersen, K., Leah, R., Knudsen, S. and Cameron-Mills, V. (2002) Matrix attachment regions (MARs) enhance transformation frequencies and reduce variance of transgene expression in barley. *Plant Mol. Biol.* **49**, 45–58.

Sengupta-Gopalan, C., Reichert, N.A., Barker, R.F., Hall, T.C. and Kemp, J.D. (1985) Developmentally regulated expression of the β-phaseolin gene in tobacco seed. *Proc. Natl Acad. Sci. USA*, **82**, 3320–3324.

Shibata, Y., Carninci, P., Watahiki, A., Shiraki, T., Konno, H., Muramatsu, M. and Hayashizaki, Y. (2001) Cloning full-length, cap-trapper-selected cDNAs by using the single-strand linker ligation method. *BioTechniques*, **30**, 1250–1254.

Short, J.M., Fernandez, J.M., Sorge, J.A. and Huse, W.D. (1988) Lambda ZAP: a bacteriophage lambda expression vector with *in vivo* excision properties. *Nucleic Acids Res.* **16**, 7583–7600.

Singh, S.P., Zhou, X.R., Liu, Q., Stymne, S. and Green, A.G. (2005) Metabolic engineering of new fatty acids in plants. *Curr. Opin. Plant Biol.* **8**, 197–203.

Smith, M.A., Moon, H., Chowrira, G. and Kunst, L. (2003) Heterologous expression of a fatty acid hydroxylase gene in developing seeds of *Arabidopsis thaliana*. *Planta*, **217**, 507–516.

Spiess, A.N. and Ivell, R. (2002) A highly efficient method for long-chain cDNA synthesis using trehalose and betaine. *Anal. Biochem.* **301**, 168–174.

Stuitje, A.R., Verbree, E.C., van der Linden, K.H., Mietkiewska, E.M., Nap, J.P. and Kneppers, T.J.A. (2003) Seed-expressed fluorescent proteins as versatile tools for easy (co)transformation and high-throughput functional genomics in *Arabidopsis*. *Plant Biotechnol. J.* **1**, 301–309.

Suh, M.C., Schultz, D.J. and Ohlrogge, J.B. (2002) What limits production of unusual monoenoic fatty acids in transgenic plants? *Planta*, **215**, 584–595.

Verdaguer, B., de Kochko, A., Beachy, R.N. and Fauquet, C. (1996) Isolation and expression in transgenic tobacco and rice plants, of the cassava vein mosaic virus (CVMV) promoter. *Plant Mol. Biol.* **31**, 1129–1139.

Weselake, R.J. (2005) Storage lipids. In *Plant Lipids – Biology, Utilization and Manipulation* (Murphy, D.J., ed.). Oxford, UK: Blackwell, pp. 162–225.

Zeng, G. (1998) Sticky-end PCR: new method for subcloning. *BioTechniques*, **25**, 206–208.

