

Comparative analysis of expressed sequence tags from *Sesamum indicum* and *Arabidopsis thaliana* developing seeds

Mi Chung Suh^{1,*}, Mi Jung Kim¹, Cheol-Goo Hur², Jung Myung Bae¹, Young In Park¹, Chung-Han Chung³, Churl-Whan Kang⁴ and John B. Ohlrogge⁵

¹Graduate School of Biotechnology, Korea University, 1,5-Ka, Anam-dong, Sungbuk-ku, Seoul 136-701, South Korea (*author for correspondence; e-mail michung@korea.ac.kr); ²National Center for Genome Information, Korea Research Institute of Bioscience and Biotechnology, Yusong, Taejeon 305-600, South Korea; ³Department of Biotechnology, Dong-A University, Pusan 604-714, South Korea; ⁴National Crop Experiment Station, RDA, Suwon 441-100, South Korea; ⁵Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA

Received 14 November 2002; accepted in revised form 23 March 2003

Key words: ESTs (expressed sequence tags), lignans, non-green oilseed, sesame, *Sesamum indicum*

Abstract

Sesame (*Sesamum indicum*) is an important oilseed crop which produces seeds with 50% oil that have a distinct flavor and contains antioxidant lignans. Because sesame lignans are known to have antioxidant and health-protecting properties, metabolic pathways for lignans have been of interest in developing sesame seeds. As an initial approach to identify genes involved in accumulation of storage products and in the biosynthesis of antioxidant lignans, 3328 expressed sequence tags (ESTs) were obtained from a cDNA library of immature seeds 5–25 days old. ESTs were clustered and analyzed by the BLASTX or FASTAX program against the GenBank NR and *Arabidopsis* proteome databases. To compare gene expression profiles during development of green and non-green seeds, a comparative analysis was carried out between developing sesame and *Arabidopsis* seed ESTs. Analyses of these two seed EST sets have helped to identify similar and different gene expression profiles during seed development, and to identify a large number of sesame seed-specific genes. In particular, we have identified EST candidates for genes possibly involved in biosynthesis of sesame lignans, sesamin and sesamol, and also suggest a possible metabolic pathway for the generation of cofactors required for synthesis of storage lipid in non-green oilseeds. Seed-specific expression of several candidate genes has been confirmed by northern blot analysis.

Abbreviations: DAD 1-like acylhydrolase, defective in anther dehiscence 1-like acylhydrolase; EST, expressed sequence tag; NCBI, National Center for Biotechnology Information; PEP, phosphoenol pyruvate; Rubisco; ribulose-biphosphate carboxylase; UTR; untranslated region

Introduction

Sesame (*Sesamum indicum* L.) belongs to the Pedaliaceae family and is an annual, self-pollinated oilseed crop. Sesame oil, comprising 50% of the seed dry weight, has been preferentially consumed in oriental

The nucleotide sequence data of developing sesame seeds reported will appear in the dbEST GenBank database under accession numbers 14128306–14131633.

food because of its distinctive flavor and the stabilizing antioxidant properties of sesame lignans. During roasting of sesame seeds, volatile sesame flavor is generated from the chemical reaction of glyco and sulfur compounds, and various pyrazines or purans, but its major component has not yet been identified. In spite of extensive efforts to develop new sesame varieties by conventional and mutational breeding, the lack of a non-shattering sesame variety is one of the

major barriers to obtaining high yield of sesame seeds (Yermanos *et al.*, 1972; Ashri, 1987). In addition, after oil extraction, the remaining meal corresponding to 50% of seed dry weight is wasted or used for feeding poultry. Therefore, identification of novel genes involved in biosynthesis of sesame-specific flavor or lignans and understanding the metabolic pathways from photosynthates toward storage oil are desirable as an aid to improve the quality and quantity of oil in sesame cultivars.

The potent antioxidant properties of sesame seed oil are attributed mainly to the presence of the lignans, sesamin and sesamolin, produced during seed development. Sesame lignans comprising 0.5–1.0% of seed dry weight have also been reported to have various biological activities including inhibiting $\Delta 5$ -desaturase in mammals, elevating plasma γ -tocopherol, enhancing vitamin E bioactivity, and inhibiting proliferation of human cancer cells (Shimizu *et al.*, 1991; Coonev, 2001; Miyahara *et al.*, 2001). Although sesame lignans have long been of interest as a dietary supplement, most of the enzymes involved in the biosynthetic pathways leading to sesame lignans have not been identified owing to the difficulty of protein purification.

Expressed sequence tags (ESTs) generated by large-scale single-pass cDNA sequencing have proven valuable for the identification of novel genes in specific metabolic pathways. As recent examples, EST projects led to identification of 'conjugase' enzymes responsible for formation of the conjugated double bonds of α -eleostearic and α -parinaric acids of *Momordica charantia* and *Impatiens balsamina* (Cahoon *et al.*, 1999) and to the identification of cyclopropane fatty acid synthase from *Sterculia foetida* seeds (Bao *et al.*, 2002). The large number of ESTs accumulating in the public database is also a potential resource to enable comparison of gene expression levels from different species, tissues, and developmental stages (Audic and Claverie, 1997; Mekhedov *et al.*, 2000). Using the *Arabidopsis* and rice expressed sequence tag (EST) sets, Ewing *et al.* (2000) compared common or different gene expression between two subclasses of the flowering plants Dicotyledonae and Monocotyledonae. Multidimensional analysis of EST data from various plant species can provide further insight not only into the functional annotation of anonymous genes, but also into identification of novel regulatory elements and representative genes involved in metabolic pathways.

Oilseed species can be divided into two groups depending on whether the seeds are green or non-green during development. Because green seeds contain photosynthetic plastids with the appearance of thylakoids, at least part of the reducing power and ATP required for the biosynthesis of storage lipid could be generated by photosynthesis. In contrast, in non-green seeds, all cofactors must be produced by catabolism of compounds provided to the seed by the mother plant (Dennis and Mierynk, 1982; Browse and Slack, 1985). Glucose-6-phosphate, pyruvate, and PEP have been proposed as major substrates for the biosynthesis of plastid fatty acids during *Arabidopsis* and rapeseed embryo development (Eastmond and Rawsthorne, 2000; Ruuska *et al.*, 2002; White *et al.*, 2002). In contrast, malate was more preferable as a substrate than pyruvate for the synthesis of fatty acid in leucoplasts isolated from castor endosperm (Eastmond *et al.*, 1997). Although import of malate to mitochondria has been known to be involved in transferring reducing equivalents from the cytosol into mitochondria, the mechanism(s) that generates or transfers reducing power in plastids is not well understood for non-green seeds.

In this study, 3328 ESTs were generated from developing sesame seeds and analyzed with BLASTX or FASTAX against NCBI and *Arabidopsis* proteome databases. This new data set has greatly increased the amount of sequence information available for sesame and the Pedaliaceae family. In addition, a comparative analysis between ESTs from developing sesame and *Arabidopsis* seeds was performed with a focus on ESTs involved in biosynthesis of sesame lignans and ESTs associated with accumulation of storage oil from photosynthate.

Materials and methods

Plant materials

Sesamum indicum cultivars were obtained from National Crop Experiment Station, Rural Development Administration, South-Korea. Plants were grown under glass at 26 ± 3 °C. Plant tissues were immediately utilized or frozen in liquid nitrogen after harvesting and stored at -70 °C until use. For harvest of seeds at defined developmental stages, seeds were tagged on the day of flower opening.

Fatty acid analysis

Each sample was transmethylated at 90 °C for 1 h in 0.5 ml toluene and 1 ml 1 N methanolic HCl. Heptadecanoic acid (17:0) was added to each sample as an internal standard. After transmethylation, 1 ml of aqueous 0.9% NaCl was added, and fatty acid methyl esters (FAMES) recovered by three sequential extractions with 1 ml of hexane were analyzed by gas chromatography on a 30 m × 0.25 mm HP-innowax column (Hewlett Packard); oven temperature was programmed from 150 °C to 220 °C at 5 °C/min. The identity of fatty acids was determined by comparing retention times and mass spectra with standards.

Construction of cDNA library and generation of ESTs

Developing sesame seeds were mainly harvested at 5 to 25 days after flowering (DAF). Total RNA was isolated from developing seeds and seedlings with TRIzol reagent (Gibco-BRL) according to the manufacturer's instruction. Poly(A)⁺ RNA was purified with PolyAtract mRNA isolation systems (Promega) and a cDNA library was constructed with a uni-ZAP cDNA synthesis kit as described in the manufacturer's instructions (Stratagene, USA). After *in vivo* mass excision of the cDNA library, more than 5000 colonies were randomly selected and cultivated for isolation of plasmid at 37 °C. Plasmid DNA was prepared by alkaline lysis (Sambrook *et al.*, 2001) and column purification (Quiagen). Automated fluorescence cycle sequencing reactions were done with an ABI 'Big dye' cycle sequencing kit (PE Applied Biosystems) with T3 (AATTAACCTCACTAAAGGG) primers. Reactions were run and analyzed on an ABI Prism 3700 DNA Analyzer. Geospiza Finch Server software (<http://www.apple.com/scitech/lifescience/genomics>) was used to screen out low-quality DNA sequences. The vector sequence was masked with Cross_match. After screening, sequences were clustered using StackPack (<http://www.sanbi.ac.za/CODES/>). Groups that contain only one sequence were classified as singletons.

Database searches

The BLASTX searches against NR, NCBI, and *Arabidopsis* proteome sequence (<http://mips.gsf.de/proj/thal/index.html>) were performed at a threshold of BLAST scores >50 or an E-value >1.0E-05. We also used FASTAX, as a more sensitive program (<http://dna.uio.no/search>) to maximize identification

of predicted function of sesame ESTs. Comparative analyses between sesame and *Arabidopsis* were carried out with the AGI locus ID of the *Arabidopsis* proteome corresponding to each EST. All information of developing sesame seed ESTs, including BLASTX analyses against NCBI and *Arabidopsis* proteome databases, will be presented at our website (<http://biotech.korea.ac.kr/member/michung/index.html>). To provide broad information for each gene in our website, the AGI locus ID of the *Arabidopsis* proteome was hyperlinked to the website MIPS (<http://mips.gsf.de/>).

Northern blot analysis

Total RNAs were isolated from leaves, 15-day old seedlings, and developing seeds of sesame with TRIzol reagent (Gibco-BRL) according to the manufacturer's instructions. About 20 µg of total RNA was loaded and transferred onto positively charged nylon membrane (Tropix). The membrane was pre-hybridized and hybridized in 0.25 M sodium phosphate pH 7.2, 7% SDS, and 1 mM EDTA with biotin-labeled probes (dCTP-biotin, Sigma) at 65 °C. After washing the membrane in 0.1× SSC and 0.1% SDS at 65 °C, the signal was visualized with Southern-Star (Tropix) according to the manufacturer's instructions.

Western blot analysis

Developing seeds of sesame or *Arabidopsis* were homogenized in 1× SDS sample loading buffer (62.5 mM Tris-HCl pH 6.8, 2% SDS, 5% glycerol, 0.1 M DTT, and 0.025% bromophenol blue) and incubated at 95 °C for 5 min. After centrifugation of the crude extract at 15 000 × g for 10 min, aliquots (10 µg) of the supernatant were loaded on 12% SDS-PAGE according to the method of Laemmli (1970). Proteins were transferred onto nitrocellulose membrane and hybridized with antiserum against alfalfa cMDH provided by Dr. Carroll P. Vance (University of Minnesota, USA) as described by Harlow and Lane (1988).

Results and discussion

Generation of 3328 ESTs from developing sesame seeds

To select a sesame cultivar for the generation of developing sesame seed ESTs, over 10 sesame cultivars

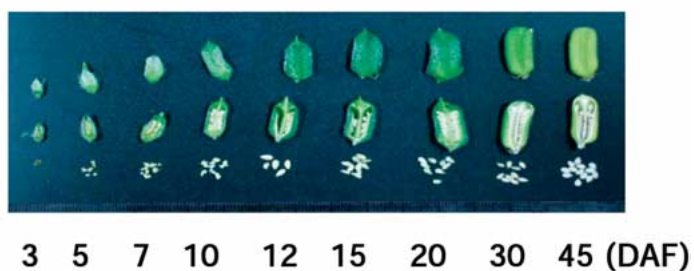
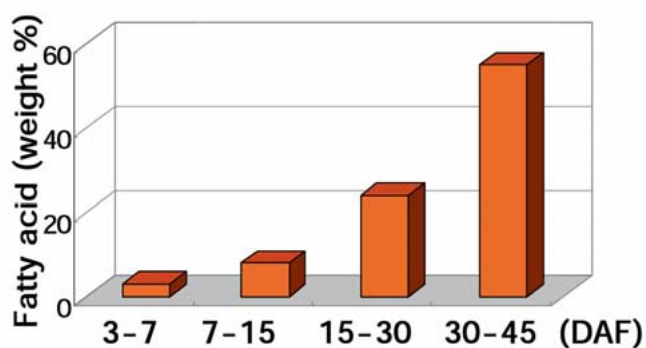
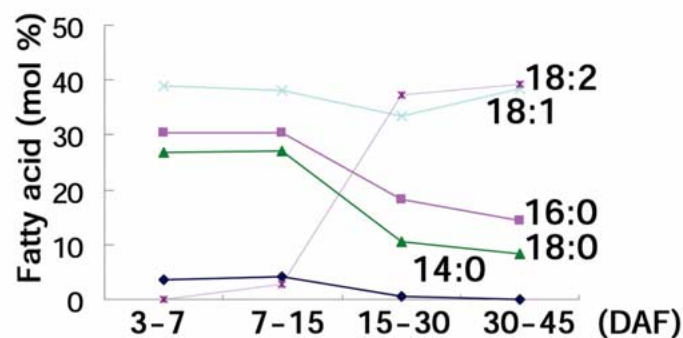
(A) Developing seeds of *S. indicum***(B) Lipid deposition****(C) Lipid composition**

Figure 1. Developmental analysis of lipid deposition and composition from sesame seeds. A. The developing seeds were tagged on the day of flower opening so that seeds could be harvested according to developmental stages. B and C. Fresh weight of developing seeds and dry weight of mature seeds were used for fatty acid analysis. Fatty acid methyl esters from seeds were analyzed by gas chromatography using an HP-innowax column (Hewlett Packard). The values are the average of three independent analyses.

currently cultivated in South Korea were analyzed for amount and composition of fatty acids. All cultivars examined showed similar oil content and composition per seed dry weight (data not shown). Considering annual productivity and oil content of sesame seeds, *Sesamum indicum* cv. Yangbaeck was selected and further analyzed for lipid deposition and composition during seed developmental stages. When developing sesame seeds were examined from less than 7 to 45 DAF, storage oil continuously accumulated and the amount of linoleic acid increased from 5% to 30% of total fatty acids between 15 and 30 DAF (Figure 1). This period encompasses seed developmental stages during which deposition and composition of storage oil rapidly changed. When sesame seeds were mature, oil accumulated up to 55% of seed dry weight. In mature sesame seeds, the amount of palmitic acid and stearic acid decreased to ca. 13% and 10% of total fatty acids, respectively, whereas total unsaturated fatty acids, (oleic acid and linoleic acid), continuously increased up to 75–80% of total fatty acids (Figure 1).

A cDNA library was constructed by using mRNA isolated from immature seeds 5–25 days old, the average time period for initiation of storage oil biosynthesis and the stage of high accumulation of sesame lignans (Jiao *et al.*, 1998). Plasmid DNAs were automatically prepared from more than 4000 cDNA clones and sequenced from the 5' end. After screening of low-quality DNA sequences and trimming of vector sequences, the sesame ESTs were clustered to classify the redundant ESTs as well as singletons. Clustering of 3328 ESTs with StackPack software generated 623 independent clusters from 2219 ESTs, and 1109 singletons. Although some ESTs represent non-overlapping clones from the same gene, these data suggest that about 50% of total sesame ESTs may represent independent sesame genes.

Analysis of sesame ESTs against NR and Arabidopsis proteome databases

The 623 consensus sequences from clusters and 1109 singletons were primarily analyzed with BLASTX against the non-redundant (NR) protein database of GenBank. About 85% of sesame ESTs had significant (E-value <1.0E–10) similarity to the translation products in the NR database at the time of analysis. About 16% of sesame ESTs, although showing significant similarity (BLAST scores >50) to proteins of the NR database, were classified as 'function unknown'. To search for novel or seed-specific genes

expressed during sesame seed development, the deduced amino acid sequences of sesame ESTs were also analyzed by comparison to the *Arabidopsis* proteome. More than 90% of all sesame ESTs was judged to have the same putative function through matches to both the NR database and the *Arabidopsis* proteome. Most sesame ESTs having no match or low similarity in the NR database showed a similar sequence homology against the *Arabidopsis* proteome. However, ca. 2% of sesame ESTs had significant 'hits' against the NR database, but showed no match or low similarity against the *Arabidopsis* proteome. For example, the biggest difference in score probability between the NR and *Arabidopsis* proteome databases was for a sesame EST cluster corresponding to 2S albumin, a seed storage protein. This finding was consistent with the report that sesame 2S albumin was completely different from *Arabidopsis* 2S albumin except for 8 conserved cysteine residues (Tai *et al.*, 1999). Interestingly, ESTs encoding a subunit of multi-component enzyme complexes such as NADH:plastoquinone oxidoreductase chain 4, mitochondrial ATPase β subunit, and chlorophyll *a,b*-binding protein cp24 10A precursor also had different score probabilities between the NR and *Arabidopsis* proteome databases. Overall, ca. 17% of sesame ESTs showed no match or low similarity against the *Arabidopsis* proteome and therefore may be novel genes in sesame related species. However, the possibility that some of these ESTs represent mainly 5'- or 3'-UTR sequences cannot be excluded.

To compare gene expression between developing sesame and *Arabidopsis* seeds, the sesame ESTs were compared to a set of ESTs from developing *Arabidopsis* seeds (White *et al.*, 2000). The comparison between developing sesame and *Arabidopsis* seed ESTs is especially interesting because these two species represent examples of plants with non-green and green seeds, respectively. After BLASTX searches of 3,328 sesame and 10,032 *Arabidopsis* ESTs against the *Arabidopsis* proteome, the gene locus ID of the *Arabidopsis* protein showing highest sequence similarity (BLAST score >50) was individually assigned to the corresponding sesame or *Arabidopsis* EST. Sesame and *Arabidopsis* ESTs annotated with the gene locus ID of the *Arabidopsis* proteome were classified into functional categories assigned by Munich Information Center for Protein Sequences (<http://mips.gsf.de>). Table 1 presents a comparison of functional assignments between sesame and *Arabidopsis* developing seed ESTs. Some ESTs were classified in two or three functional categories. ESTs assigned to the categor-

Table 1. Functional catalog of sesame and *Arabidopsis* developing seed ESTs. Functional assignment of developing sesame and *Arabidopsis* seed ESTs was done by sequence homology to *Arabidopsis* proteins in the MIPS database. Dual or triple functions were assigned to some ESTs.

Functional category	Description	Sesame (% of ESTs)	<i>Arabidopsis</i> (% of ESTs)
1	metabolism	27.0	40.0
2	energy	7.4	12.9
3	cell growth, cell division and DNA synthesis	7.2	9.3
4	transcription	7.0	12.0
5	protein synthesis	9.8	5.9
6	protein destination	17.6	29.0
7	transport facilitation	4.0	4.1
8	cellular transport and transport mechanisms	7.1	7.2
9	cellular biogenesis	3.3	5.0
10	cellular communication/signal transduction	5.2	6.5
11	cell rescue, defense, cell death and aging	13.0	17.5
13	ionic homeostasis	0	0
30	cellular organization	36.8	41.5
35	motility	0	0
50	development	0	0.3
90	transposable elements, viral and plasmid proteins	0.6	0.2
97	organism-specific proteins	0	0
98	classification not yet clear	13.2	13.3
99	unclassified proteins	22.0	30.7

ies of metabolism and cellular organization (proteins localized to the corresponding organelles) were predominant in both developing sesame and *Arabidopsis* seeds. More than 30% of sesame ESTs were in the category of unclassified proteins or undetermined classification. Overall the ratios of ESTs classified to each functional category were similar between sesame and *Arabidopsis*, except for the category of protein synthesis, where the percentage of sesame ESTs was higher than that of *Arabidopsis* ESTs. In particular, the number of ESTs encoding ribosomal proteins assigned in the protein synthesis category was about six-fold higher in sesame than *Arabidopsis*. The number of ESTs encoding 12 seed storage proteins in the protein destination category was 3.6-fold higher in *Arabidopsis* than in sesame, even though some cDNA clones encoding seed storage proteins were subtracted before sequencing of *Arabidopsis* seed ESTs (White *et al.*, 2000). In combination with reports that transcripts of 12S seed storage protein increased 10-fold in mature seeds of *Arabidopsis* (Ruuska *et al.*, 2002) and transcripts of ribosomal protein were abundant in cell division or proliferation (Van Lijsebettens *et al.*, 1994), this result suggested that the seed develop-

mental stage of sesame might be earlier than that of *Arabidopsis* used for the construction of the cDNA library.

Comparative analysis of most abundant ESTs of sesame and Arabidopsis developing seeds

To investigate similarities or differences in gene expression between sesame and *Arabidopsis* seeds, the two seed EST sets were compared. To help provide a common reference for this comparison, sesame ESTs (both contigs and singletons) were analyzed by BLASTX against the *Arabidopsis* proteome and assigned the AGI locus ID of the closest *Arabidopsis* protein. However, because many *Arabidopsis* genes occur as gene families, individual inspections and grouping of genes were conducted based on annotation or description of each gene or enzyme reaction.

Table 2 presents a comparison of genes that are highly expressed in sesame seeds (based on EST numbers) with those of *Arabidopsis*. The number of sesame or *Arabidopsis* ESTs associated with each gene locus ID of the *Arabidopsis* proteome provides an estimate of the expression level of the respective gene.

Table 2. Comparison of most abundant seed ESTs of sesame with *Arabidopsis* seed ESTs. Developing sesame and *Arabidopsis* seed ESTs were analyzed by BLASTX against *Arabidopsis* proteome sequences. Sesame contigs with more than ten ESTs were compared with developing *Arabidopsis* seed ESTs by AGI locus ID of the *Arabidopsis* proteome.

FC ^a	Query ^b	Sesame (ESTs)	<i>Arabidopsis</i> (ESTs)	Hit ID of <i>Arabidopsis</i>	Description	E-value
UC	cl15ct28cn37	68	–	At1g70850	major latex protein (MLP149)	3.63E–43
		–	2	At1g14950	major latex protein type1	
	cl40ct49cn60	51	–	At3g09390	metallothionein-like protein	1.79E–10
	cl28ct42cn52	64	–	no hits found	putative allergen protein	3.00E–15
	cl319ct334cn360	10	–	At4g38400	putative pollen allergen	8.38E–52
	cl142ct151cn171	9	–	At1g29140	hypothetical protein similar to allergen	2.00E–26
	cl142ct150cn170	2	2	At1g78040	allergen, putative similar to allergen	8.21E–10
		–	1	At4g12730	putative pollen surface protein	
		–	1	At4g36750	minor allergen	
	cl52ct58cn70	24	–	no hits found	hydroxyproline-rich protein	7.00E–20
	cl43ct48cn58	18	–	At1g12090	aluminum induced gene	5.26E–18
	cl174ct184cn205	12	–	At5g22580	unknown protein [<i>A. thaliana</i>]	8.77E–26
	cl89ct96cn113	11	1	At5g55410	unknown protein [<i>A. thaliana</i>]	1.76E–15
	cl1ct9cn11	52	83	At1g03880	cruciferin 12S seed storage protein	2.56E–65
SP	cl35ct38cn48	19	40	At1g03890	cruciferin 12S seed storage protein	1.39E–79
	cl207ct218cn239	2	651	At4g28520	12S cruciferin seed storage protein	1.42E–48
	cl65ct79cn95	44	–	no hits found	2S albumin 1 (<i>Sesamum indicum</i>)	4.84E–35
	cl187ct199cn220	25	–	no hits found	2S albumin 2 (<i>Sesamum indicum</i>)	5.83E–56
	cl114ct122cn140	13	–	no hits found	2S albumin 3 (<i>Sesamum indicum</i>)	1.97E–28
		–	207	At4g27160	2S albumin 3 precursor	
		–	69	At4g27150	2S albumin 2 precursor	
		–	28	At4g27140	2S albumin 1 precursor	
		–	15	At4g27170	2S albumin 4 precursor	
	cl57ct65cn81	30	354	At5g44120	legumin-like protein [<i>A. thaliana</i>]	2.50E–101
	cl8ct15cn17	39	–	At3g18280	lipid transfer protein	9.71E–19
	cl14ct17cn24	26	–	At5g01870	lipid-transfer protein	4.09E–23
	cl97ct103cn120	11	–	At2g38540	putative non-specific lipid-transfer protein	9.33E–20
		–	12	At2g38530	putative non-specific lipid-transfer protein	
		–	5	At1g05450	lipid transfer protein, putative	
LM		–	3	At3g43720	lipid transfer protein-like protein	
		–	1	At5g38160	lipid transfer-like protein	
		–	1	At3g51600	non-specific lipid transfer protein	
	cl37ct40cn50	14	10	At4g25140	oleosin, 18.5K [<i>A. thaliana</i>]	5.65E–19
	cl139ct146cn165	11	15	At3g01570	putative oleosin	2.13E–29
		–	53	At3g27660	oleosin isoform	
		–	40	At5g40420	oleosin	
	cl92ct98cn115	16	–	At3g27890	NADPH:quinone oxidoreductase	2.01E–50
	cl398ct412cn443	3	1	At5g16990	quinone oxidoreductase-like protein	7.73E–51
	cl157ct168cn189	13	–	At1g26360	putative α/β hydrolase	7.32E–26
		–	2	At3g29770	α/β hydrolase, putative	
	cl237ct250cn271	11	–	At5g43330	cytosolic malate dehydrogenase	5.56E–104

Table 2. Continued.

FC ^a	Query ^b	Sesame (ESTs)	<i>Arabidopsis</i> (ESTs)	Hit ID of <i>Arabidopsis</i>	Description	E-value
RBP	cl528ct542cn579	13	55	At5g60390	translation elongation factor eEF-1 α chain	2.92E-119
	cl120ct127cn145	12	1	At2g21660	glycine-rich RNA binding protein 7	7.32E-37
	cl412ct426cn457	7	–	At4g13850	glycine-rich RNA-binding protein AtGRP2	3.78E-32
		–	1	At4g39260	glycine-rich protein (clone AtGRP8)	
PA	cl38ct43cn53	11	–	At1g33140	ribosomal protein L9	6.81E-75
	cl54ct59cn73	10	4	At5g50260	cysteine endopeptidase	7.74E-102
	cl546ct560cn598	8	7	At4g05320	polyubiquitin (ubq10)	6.89E-63
	cl211ct223cn244	3	1	At4g02890	polyubiquitin	2.41E-74
	MC01008C02	1	1	At5g03240	polyubiquitin (ubq3)	2.82E-44
ST		–	1	At4g12570	polyubiquitin-like protein	
	cl81ct87cn103	8	–	At5g55190	small Ras-like GTP-binding protein	5.15E-114
	MC01032C05	1	1	At4g18800	Ras-like GTP-binding protein	2.04E-85
	MC01028E09	1	4	At4g17530	Ras-like GTP-binding protein	1.15E-54
		–	2	At2g31680	RAS superfamily GTP-binding protein	
DV		–	1	At2g43130	Ras-related GTP-binding protein (ARA-4)	
	cl206ct219cn240	11	–	At1g54410	water stress-induced protein	2.14E-04

^aFC, functional category; UC, unclassified; SP, storage protein; LM, lipid metabolism; SM, secondary metabolism; CM, carbohydrate metabolism; RBP, RNA-binding proteins; PA, proteinases; ST, signal transduction; DV, development.

^bQuery refers to the identification number of sesame ESTs after clustering.

‘–’: No EST was observed in developing sesame or *Arabidopsis* seed ESTs. Two clusters, cl28ct42cn52 and cl52ct58cn70, were matched by using the FASTAX instead of the BLASTX program. Three clusters, cl65ct79cn95, cl187ct199cn220, and cl114ct122cn140, were matched by BLASTX against the NR database.

As in *Arabidopsis* developing seed ESTs (White *et al.*, 2000), the most abundant sesame ESTs were seed storage proteins, including 12S cruciferin (previously identified as 11S globulin), 2S albumin, and legumin-like protein. The number of ESTs encoding oleosin or translation elongation factor (eEF-1 α chain) was also high in both EST sets.

In contrast, ESTs corresponding to several proteins were much more abundant in sesame than in *Arabidopsis*. For example, metallothionein-like protein (51 sesame ESTs), hydroxyproline-rich protein (24 sesame ESTs), and aluminum-induced gene (18 sesame ESTs) contrast strongly with the complete lack of ESTs corresponding to these proteins from the *Arabidopsis* seed collection. Similarly, major latex protein (68 sesame ESTs) was far more abundant in sesame than *Arabidopsis* (2 ESTs).

Another abundant EST group of sesame seeds was ‘allergen-like’ proteins which cause hypersensitive response in man. Allergic reactions induced by ingestion of foods containing sesame seeds have been well documented, but allergens of sesame seeds have been little investigated. Recently, 2S albumins and lipid transfer proteins (LTPs) were purified as major allergens of sesame seeds (Pastorello *et al.*, 2001). The mRNAs encoding these two proteins comprised up to 5% of total sesame ESTs. In addition, four kinds of allergen-like proteins, cl28ct42cn52, At4g38400, At1g29140, and At1g78040, were also abundant in sesame seed ESTs (Table 2). This information may be important not only for diagnosis of food allergy, but also for production of non-allergenic drugs and cosmetics.

ESTs encoding chlorophyll *a,b*-binding proteins were more than 4-fold higher in *Arabidopsis* than in sesame (<http://biotech.korea.ac.kr/member/michung/>

index.html). The low frequency of ESTs encoding chlorophyll *a,b*-binding proteins in sesame ESTs is consistent with the fact that sesame seeds contain non-photosynthetic plastids. Interestingly, expression of Rubisco small subunits was very similar in both EST sets. Ruuska *et al.* (2002) reported that high expression of Rubisco small subunits in developing *Arabidopsis* seeds was possibly related to the recapture of CO₂ released from the conversion of pyruvate to acetyl-CoA by the plastid pyruvate dehydrogenase. It remains to be investigated whether the expression of Rubisco small subunits is also involved in CO₂ recapture in non-green seeds such as sesame.

The regulation of gene expression is mostly controlled by internal (e.g. hormones) or external (e.g. light and temperature) signals via signal receptor and signal transduction system. ESTs corresponding to small Ras-like GTP-binding protein (At5g55190, 8 sesame ESTs), calmodulin (At3g43810, 7 sesame ESTs), ADP-ribosylation factor (At2g47170, 6 sesame ESTs), or calcium-dependent protein kinase (At3g57530, 6 sesame ESTs) were about 20-fold higher in sesame than in *Arabidopsis* (<http://biotech.korea.ac.kr/member/michung/index.html>). ADP ribosylation factor is known to have an important role in vesicle transport and in the phosphoinositide-mediated signal transduction pathway (Martin, 1998; Chavrier and Goud, 1999; McElver *et al.*, 2000). Asano *et al.* (2002) reported that a calmodulin-like protein kinase specifically expressed in the endosperm of immature seed is required for accumulation of storage products during rice seed development and involved in phosphorylation of sucrose synthase. To further examine the mRNA level of genes corresponding to sesame ESTs involved in cellular signal transduction, we analyzed more than one hundred sesame ESTs by reverse northern blot. Consistent with the EST abundance, the expression of genes corresponding to ESTs encoding small Ras-like GTP-binding protein, calmodulin, and calcium-dependent protein kinase were 3- to 10-fold higher in developing seeds than in 15-day old seedlings of sesame (data not shown).

Table 2 also indicates that in several cases, although related ESTs occurred in both the sesame and *Arabidopsis* seed data sets, these ESTs mapped to different members of *Arabidopsis* multigene families. Thus, for example, the most abundant sesame ESTs for lipid transfer protein and oleosin mapped to genes At3g18280 and At4g25140, respectively, whereas the

Arabidopsis ESTs mapped to different members of the *Arabidopsis* gene family.

Comparative analysis of developing sesame and Arabidopsis seed ESTs involved in lignan biosynthesis

Although there are several limitations to the use of EST frequencies to infer metabolic or gene expression patterns in different species or different developmental stages (Audic and Claverie, 1997), consideration of such data can provide first insights and valuable clues into the sequences and expression patterns responsible for biosynthetic characteristics. Therefore, in the following sections, selected aspects of the sesame EST data set are presented and compared to *Arabidopsis* seed ESTs.

To identify EST candidates involved in the biosynthesis of sesame lignans during seed development, a biosynthetic pathway was constructed as shown in Figure 2 based on recent reports by radiolabeled and stable-isotope-labeled precursor administration experiments (Jiao *et al.*, 1998; Kato *et al.*, 1998). Tyrosine or phenylalanine is first converted into coniferyl alcohol by the phenylpropanoid pathway. Two coniferyl alcohols are subsequently coupled to pinoresinol, and then pinoresinol is converted into 'oxygen-inserted' lignans, sesamol or sesamin via piperitol containing a single methylenedioxy bridge. Preliminary identification of EST candidates corresponding to each enzymatic reaction in the biosynthetic pathway of sesame lignans are presented in Table 3 and compared with *Arabidopsis* seed ESTs.

Interestingly, no EST encoding phenylalanine ammonia-lyase was detected in the sesame seed ESTs, even though it was abundantly expressed in *Arabidopsis* seed ESTs (reactions 1 and 2 in Figure 2 and Table 3). This result supports previous biochemical results in which [U-¹⁴C] phenylalanine was not incorporated into sesamol and sesamin in developing sesame seeds, but instead [1-¹⁴C]tyrosine was converted into sesamin in cell suspension cultures of sesame (Kato *et al.*, 1998). In the phenylpropanoid pathway for the formation of coniferyl alcohol from phenylalanine or tyrosine, caffeic acid *O*-methyltransferase (COMT), which is proposed as an alternative catalyst for *O*-methylation leading to coniferyl alcohol, than caffeoyl-CoA *O*-methyltransferase (CCOMT) was predominant in developing sesame seeds (20 ESTs for reaction 5 in Figure 2 and Table 3). In contrast, in cell suspension cultures of *Pinus taeda*

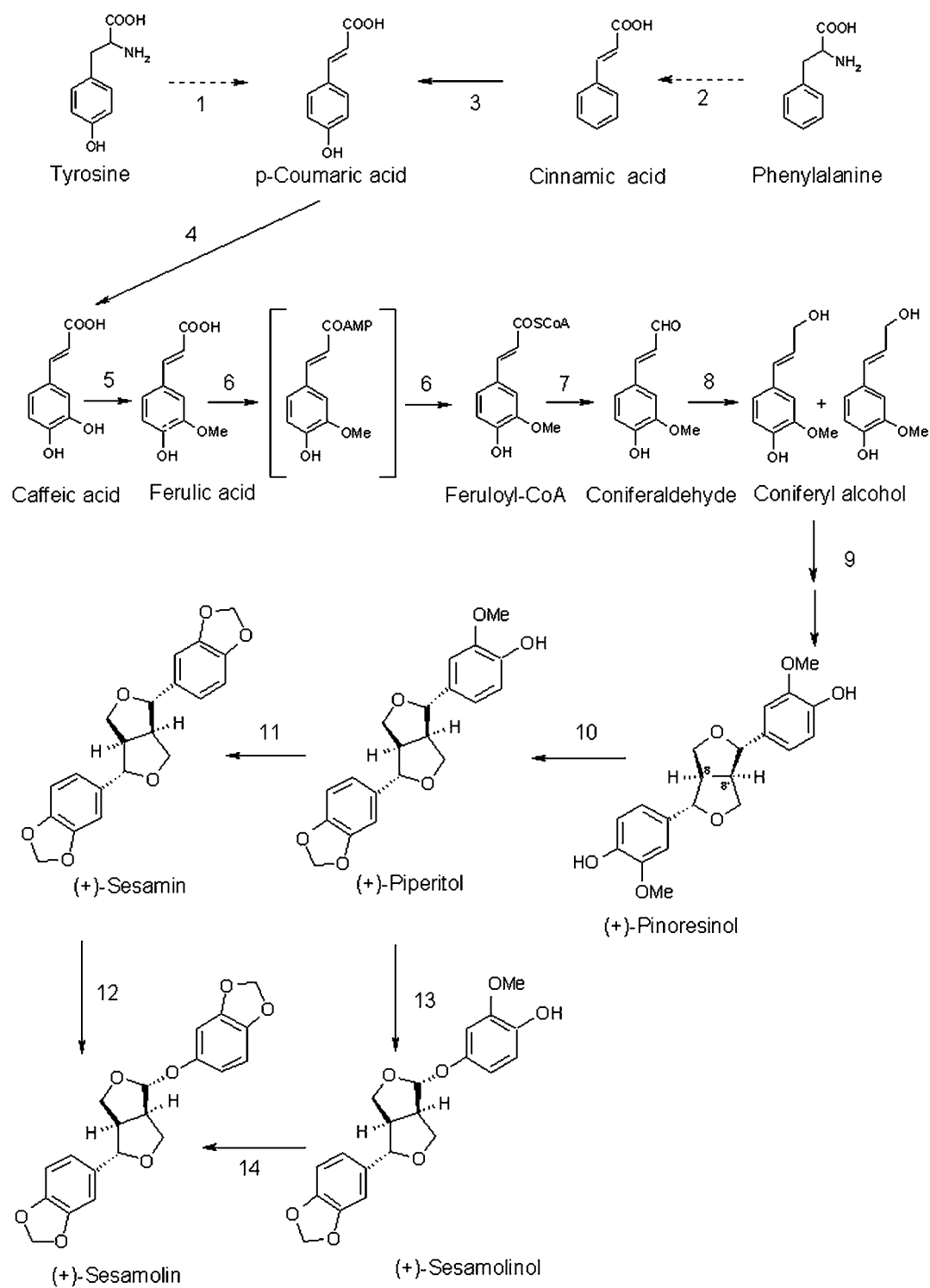


Figure 2. Possible biosynthetic pathways for sesame lignans, sesamin and sesamol. Numbers correspond to individual reactions and serve to identify the respective enzyme in Table 3. If no candidate gene responsible for each reaction was found in current sesame seed ESTs, the reaction is shown as a dashed line. Me means methyl group ($-\text{CH}_3$).

Table 3. Comparative analysis of developing sesame and *Arabidopsis* seed ESTs possibly involved in lignan biosynthesis.

No. ^a	Hit ID of <i>Arabidopsis</i>	Query ^b	Sesame (ESTs)	<i>Arabidopsis</i> (ESTs)	Description	E-value
1	–	–	–	–	tyrosine ammonia-lyase	–
2	At3g10340	–	–	9	putative phenylalanine ammonia-lyase	–
	At3g53260	–	–	4	phenylalanine ammonia-lyase	–
3	At2g30490	cl291ct304cn328	3	1	cinnamate-4-hydroxylase	2.23E–66
		MC01048A10	1			2.19E–40
4	At5g07990	MC01017D02	1	1	flavonoid-3-hydroxylase	7.90E–26
	At2g40890	cl590ct604cn643	2	1	coumaric acid hydroxylase	2.72E–29
5	At5g54160	cl25ct29cn38	5	3	caffeic acid <i>O</i> -methyltransferase	8.07E–72
		cl55ct60cn75	5			2.38E–65
		cl470ct484cn518	2			5.31E–60
	At4g35160	cl42ct46cn56	9	1	caffeic acid <i>O</i> -methyltransferase	2.62E–29
		MC01025D08	1			3.03E–18
	At3g53140	MC01042E11	1	–	caffeic acid <i>O</i> -methyltransferase	3.30E–30
	At1g77520		–	3	caffeic acid 3- <i>O</i> -methyltransferase	
	At1g63140		–	2	caffeic acid <i>O</i> -methyltransferase	
	At4g34050		–	1	caffeoyl-CoA <i>O</i> -methyltransferase	
	At4g26220		–	1	caffeoyl-CoA <i>O</i> -methyltransferase	
6	At3g16910	cl20ct22cn30	2	1	AMP-binding protein (CoA ligase)	2.38E–23
7	At1g80820	MC01037A11	1	–	cinnamoyl CoA reductase	2.74E–08
8	At4g37990	MC01037C02	1	–	cinnamyl-alcohol dehydrogenase ELI3-2	6.60E–67
	At4g37980	MC01021A06	1	–	cinnamyl-alcohol dehydrogenase ELI3-1	2.70E–39
	At5g19440	MC01045B07	1	–	cinnamyl-alcohol dehydrogenase	3.30E–80
	At1g72680	MC01032F07	1	2	putative cinnamyl-alcohol dehydrogenase	2.96E–63
	At2g21890	MC01053D07	1	–	putative cinnamyl-alcohol dehydrogenase	3.32E–30
9	At1g58170	MC01032H05	1	–	dirigent protein	2.73E–59
		cl416ct422cn426	2		dirigent protein	2.20E–20
10	At3g26200	cl606ct620cn661	2	–	cytochrome P450	2.89E–30
14	At1g11680	MC01012H10	1	3	cytochrome P450 family	4.52E–48
	At5g25900	MC01037G04	1	1	cytochrome P450	1.14E–18
	At4g24520	MC01034G01	1		NADPH-cytochrome P450 oxidoreductase	5.55E–40
	At3g14690	MC01033E07	1	–	putative cytochrome P450	3.76E–41
	At2g45550	MC01057G05	1	–	putative cytochrome P450	1.28E–13

Table 3. Continued.

No. ^a	Hit ID of <i>Arabidopsis</i>	Query ^b	Sesame (ESTs)	<i>Arabidopsis</i> (ESTs)	Description	E-value
	At4g15110	MC01053B07	1	–	cytochrome P450 protein	2.02E–55
	At4g39230	cl480ct494cn529	3	–	NAD(P)H oxidoreductase	1.73E–45
	At3g24200	cl607ct621cn662	2	–	monooxygenase	2.18E–16
	At1g52810	cl236ct249cn270	2	–	putative oxidoreductase	8.96E–07
	At1g75280	MC01032C01	1	2	NADPH oxidoreductase	3.12E–49
	At3g12870	MC01012B02	1	–	putative oxidoreductases	2.22E–14
	At1g06640	–	–	1	oxidoreductase, putative	

^aNumbers correspond to individual reactions and serve to identify the respective enzyme in Figure 2.

^bQuery refers to the identification number of sesame ESTs after clustering.

‘–’: No EST was observed in developing sesame or *Arabidopsis* seed ESTs. One cluster, cl416ct422cn426, was matched by using the FASTAX program against the NR database.

which produce lignin from coumaryl alcohol or coniferyl alcohol, the transcript level of COMT was much lower than that of CCOMT and was not up-regulated by exogenous feeding of phenylalanine (Anterola *et al.*, 2002). Thus, these EST data provide preliminary information that higher expression of COMT in developing sesame seeds might be needed for the formation of the coniferyl alcohol precursor of sesame lignans than CCOMT, involved in the production of both coumaryl alcohol and coniferyl alcohol.

It has been reported that formation of pinorelinol from coniferyl alcohol is mediated by a dirigent (Latin: guiding or aligning) protein in *Forsythia* sp. (Davin *et al.*, 1997; Burlat *et al.*, 2001). According to Kato *et al.* (1998), *E*-[8-¹⁴C]coniferyl alcohol was also metabolized into pinorelinol, piperitol, sesamin, and sesamol in developing sesame seeds. In the sesame ESTs, three ESTs encoding dirigent protein were detected (reaction 9 in Figure 2 and Table 3). This finding suggests that a dirigent protein might be involved in the reaction of bimolecular phenoxyl radicals for the formation of pinorelinol in developing sesame seeds. Recently, a methylenedioxy bridge formation for the synthesis of piperitol was reported to occur via an O₂/NADPH-requiring cytochrome P450-dependent enzyme in microsomal fractions of sesame seeds (Jiao *et al.*, 1998). However, little information is available about proteins or genes leading to sesamin or sesamol from piperitol (Marchand *et al.*, 1997). Based on the results of Jiao *et al.* (1998), sesame EST candidates possibly involved in methylenedioxy bridge formation and oxygen insertion are shown in Table 3 (reactions 10–14).

To further analyze the expression of some EST candidates involved in the biosynthesis of sesamin

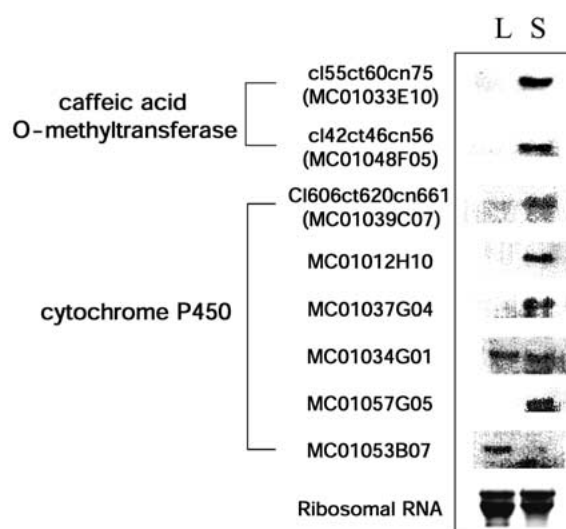


Figure 3. Northern blot analysis of EST candidates corresponding to genes involved in biosynthesis of sesame lignans. About 20 µg of total RNA isolated from sesame leaves (L) and developing seeds (S) was loaded in each lane. Ethidium bromide staining of ribosomal RNAs indicates similar loading of RNA in both lanes.

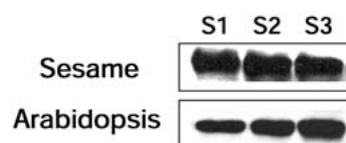


Figure 4. Western blot analysis for cytosolic malate dehydrogenase in developing sesame and *Arabidopsis* seeds. Total proteins (10 µg) isolated from developing seeds of sesame (S1, <10 DAF; S2, 11–25 DAF; S3, >25DAF) or *Arabidopsis* (S1, <7 DAF; S2, 8–15 DAF; S3, >16 DAF) were loaded on 12% SDS-PAGE and probed with antiserum against alfalfa cMDH.

and sesamol, northern blot analysis was carried out using total RNAs extracted from leaves and seeds of sesame. As shown in Figure 3, two abundant clusters, cl55ct60cn75 and cl42ct46cn56, corresponding to COMT were specifically expressed in seed tissues. Expression of some ESTs responsible for putative cytochromes P450, MC01039C07, MC01012H10, MC01037G04, and MC01053B07, was substantially higher in seeds than in leaves. In contrast, expression of the MC01053B07 clone was predominant in leaves rather than in seeds. The MC01034G01 clone encoding NADPH-cytochrome P450 oxidoreductase was expressed in both leaf and seed tissues. Because only a few genes involved in the biosynthesis of sesame lignans have been identified to date, this information may be helpful for further studies to elucidate the pathways involved in sesame lignan production.

Comparative analysis of ESTs involved in metabolic pathways responsible for accumulation of storage oil from photosynthate between sesame and Arabidopsis

During *Arabidopsis* and rapeseed seed development, the conversion of photosynthate into seed oil and its regulatory mechanism have been studied by biochemical, EST analysis and microarray approaches (Eastmond and Rawthorne, 2000; Girke *et al.*, 2000; White *et al.*, 2000; Ruuska *et al.*, 2002; Schwender and Ohlrogge, 2002). However, comparatively little information is available for non-green seeds such as sunflower, cotton, or sesame. As a first step toward understanding differences in metabolic pathways responsible for the accumulation of storage oil from photosynthate between non-green seeds and green seeds, we performed a comparative analysis of sesame and *Arabidopsis* seed ESTs. From this comparison, Table 4 presents enzymes or proteins with more than a 3-fold difference in EST numbers between developing sesame and *Arabidopsis* seeds. Differences that are statistically significant at the 95% confidence level are underlined.

Considering that the total number of sesame ESTs was about 3-fold smaller than that of *Arabidopsis* ESTs, one striking difference was that 86 ESTs encoding lipid transfer protein (LTP) were generated from 3328 sesame ESTs. Two major LTP isoforms (homologous to At3g18280 and At5g01870) comprise ca. 2% of total sesame ESTs. Although 20 genetic loci encoding LTPs are found in the *Arabidopsis* genome, the transcript level of LTPs was 10-fold lower in *Arabidopsis* seeds than in sesame. Recently, 130 ESTs

encoding LTP were isolated from a total of 1794 rose petal ESTs (Channeliere *et al.*, 2002). Plant LTPs have been suggested to play a role in cutin formation, embryogenesis, defense reactions against phytopathogens, and as an allergenic molecule, but the exact function of LTP is still unknown (Kader, 1996). Like LTP, ESTs encoding phosphatidylinositol transfer protein (PITP) are much higher in sesame than in *Arabidopsis*. Phosphatidylinositol transfer protein was originally defined by its ability to transfer phosphatidylinositol or phosphatidylcholine monomers between bilayers *in vitro* (Wirtz, 1991). Soybean PITPs were reported to be involved in cellular signaling to alter the plant capacity to synthesize phosphoinositides during periods of hyperosmotic stress (Monks *et al.*, 2001). Further study of abundant LTP and PITP in developing sesame seeds therefore may be of particular interest.

One interesting aspect from the comparison of sesame and *Arabidopsis* was that the number of ESTs encoding plastidic pyruvate kinase (At3g22960, At5g52920, and At3g52990) was ca. 7-fold higher in *Arabidopsis* than in sesame (Table 4). In contrast, the number of ESTs corresponding to cytosolic malate dehydrogenase (cMDH, At5g43330) was more than 30-fold higher in sesame than in *Arabidopsis* (Table 4). In developing *Arabidopsis* seeds, high expression of plastidic pyruvate kinase in conjunction with abundance of the phosphoenolpyruvate (PEP) transporter suggested a major route of carbon from sucrose into precursors of fatty acid biosynthesis involving the cytosolic glycolytic pathway up to PEP, import of this compound into the plastid, and subsequent conversion to pyruvate (White *et al.*, 2000; Ruuska *et al.*, 2002). However, higher expression of cytosolic malate dehydrogenase and lower expression of plastidic pyruvate kinase in developing seeds of sesame than in *Arabidopsis* may suggest the presence of a different metabolic pathway to generate reducing power as well as carbon source for storage oil synthesis. It has been well known that importing of malate to mitochondria was involved in the transfer of reducing equivalents from cytosol into the mitochondria. Therefore, it is possible that malate produced by reduction of oxaloacetate via cytosolic malate dehydrogenase is imported into plastids, and converted into pyruvate and CO₂, with the concomitant reduction of NADP⁺ to NADPH by plastidic NADP-dependent malic enzyme in developing sesame seeds. The generated pyruvate could be a substrate of pyruvate dehydrogenase to produce acetyl-CoA for fatty acid biosynthesis. This explanation is consistent with some biochemical evidence

Table 4. Comparative analysis of some developing sesame and *Arabidopsis* seed ESTs involved in metabolic pathways for conversion of photosynthate into seed oil.

Hit ID of <i>Arabidopsis</i>	Query ^a	Sesame (ESTs)	<i>Arabidopsis</i> (ESTs)	Statistic ^b	Description	E-value
At5g20830	MC01009G07	1	0	0.850	sucrose synthase, cytosolic	6.42E-13
At5g49190	–	–	12	<u>0.955</u>	sucrose synthase, cytosolic	–
At5g17310	cl50ct55cn66	7	3	<u>0.997</u>	UDP-glucose pyrophosphorylase, cytosolic	1.04E-120
At3g03250	–	–	2	0.150	putative UDP-glucose pyrophosphorylase, cytosolic	–
At1g79550	cl424ct438cn469	4	1	<u>0.990</u>	phosphoglycerate kinase, cytosolic	1.17E-80
At5g43330	cl237ct250cn271	11	–	<u>≥0.999</u>	cytosolic malate dehydrogenase	5.56E-104
At5g54800	–	–	4	0.550	glucose-6-phosphate phosphate translocator, plastidic	–
At1g61800	–	–	3	0.350	glucose-6-phosphate phosphate translocator, plastidic	–
At2g01140	–	–	20	<u>0.995</u>	putative aldolase, plastidic	–
At4g38970	–	–	13	<u>0.960</u>	putative fructose-bisphosphate aldolase, plastidic	–
At4g10750	–	–	1	0.150	putative aldolase, plastidic	–
At3g22960	MC01018E12	1	26	<u>0.993</u>	pyruvate kinase, putative, plastidic	4.65E-49
At5g52920	MC01002F07	1	16	0.915	pyruvate kinase, plastidic	6.34E-17
At3g52990	–	–	9	0.850	pyruvate kinase-like protein, plastidic	–
At1g01090	cl308ct321cn345	3	14	0.350	pyruvate dehydrogenase E1 α subunit, plastidic	1.02E-24
At1g59900	–	–	8	0.850	pyruvate dehydrogenase E1 α subunit, plastidic	–
At1g30120	–	–	8	0.850	pyruvate dehydrogenase E1 beta subunit, plastidic	–
At1g01630	MC01046D07	1	–	0.850	polyphosphoinositide-binding protein	3.60E-54
At1g55840	cl125ct132cn150	6	3	<u>0.993</u>	polyphosphoinositide-binding protein	2.77E-69
At1g77420	MC01024B03	1	–	0.850	monoacylglycerol lipase, mitochondrial	1.70E-18
At3g15650	MC01046F04	8	–	<u>≥0.999</u>	putative lysophospholipase, mitochondrial	4.19E-26
At5g46290	MC01056A06	1	12	0.750	3-oxoacyl-ACP-synthase I, plastidic	2.62E-55
At1g74960	MC01031D03	1	2	0.450	3-oxoacyl-ACP-synthase I, plastidic	6.92E-04
At1g24360	MC01055G05	1	8	0.550	3-oxoacyl-ACP-synthase I, plastidic	1.02E-28
At1g43800	MC01040E09	1	–	0.850	stearoyl-ACP desaturase	1.92E-62
At2g43710	–	–	10	0.915	stearoyl-ACP desaturase, plastidic	–
At3g02630	–	–	6	0.750	stearoyl-ACP desaturase, plastidic	–
At3g02610	–	–	1	0.150	stearoyl-ACP desaturase, plastidic	–
At4g33030	MC01044C09	1	1	0.650	sulfolipid biosynthesis protein, plastidic	2.38E-26
At5g01220	cl371ct385cn415	3	1	<u>0.965</u>	sulfolipid synthase, plastidic	2.86E-39
At4g30950	MC01017C11	1	–	0.850	ω -6 fatty acid desaturase, plastidic	4.79E-56
At3g12120	–	–	23	<u>0.997</u>	ω -6 fatty acid desaturase, ER	–
At2g30550	cl315ct328cn353	6	–	<u>≥0.999</u>	DAD1-like acylhydrolase	1.75E-54
At3g48460	cl362ct376cn405	3	1	<u>0.965</u>	DAD1-like acylhydrolase	1.06E-18

Table 4. Continued.

Hit ID of <i>Arabidopsis</i>	Query ^a	Sesame (ESTs)	<i>Arabidopsis</i> (ESTs)	Statistic ^b	Description	E-value
At1g09390		–	3	0.350	DAD1-like acylhydrolase	–
At1g28640		–	2	0.150	DAD1-like acylhydrolase	–
At1g28600		–	1	0.150	DAD1-like acylhydrolase	–
At1g68530	MC01056D04	1	8	0.550	beta-ketoacyl-CoA synthase	4.50E–86
At1g25450		–	1	0.150	beta-ketoacyl-CoA synthase	–
At2g26250	MC01035C04	1	14	0.850	beta-ketoacyl-CoA synthase	2.97E–37
At2g26640		–	3	0.350	beta-ketoacyl-CoA synthase	–
At3g18280	cl8ct15cn18	39	–	<u>> 0.999</u>	lipid transfer protein type2	3.52E–13
At5g01870	cl14ct17cn24	26	–	<u>> 0.999</u>	lipid transfer protein type1	4.09E–23
At2g38540	cl97ct103cn120	11	–	<u>> 0.999</u>	lipid transfer protein type1	9.33E–20
At3g51590	cl93ct99cn116	3	–	<u>0.992</u>	lipid transfer protein type1	1.08E–19
At5g59320	MC01020D06	2	–	<u>0.965</u>	lipid transfer protein type1	3.54E–12
At1g27950	cl587ct601cn640	2	–	<u>0.965</u>	lipid transfer protein	1.14E–17
At1g70240	MC01054D07	1	–	0.850	lipid transfer protein	5.70E–10
At1g73560	MC01049H06	1	–	0.850	lipid transfer protein	1.64E–13
At5g57800	MC01041H04	1	–	0.850	lipid transfer protein	1.33E–18
At3g43720		–	3	0.350	lipid transfer protein type1	
At3g51600		–	1	0.150	lipid transfer protein type1	
At2g38530		–	12	<u>0.955</u>	lipid transfer protein type1	
At1g05450		–	5	0.650	lipid-transfer protein, putative	
At5g38160		–	1	0.150	lipid transfer-like protein	

^aQuery refers to the identification number of sesame ESTs after clustering.

‘–’: No EST was observed in developing sesame or *Arabidopsis* seed ESTs. Assignment of the plastidic and cytosolic isoforms was obtained from Munich Information Center for Protein Sequences.

^bStatistic indicates the significance of differential transcript levels of genes between developing sesame and *Arabidopsis* ESTs (<http://igs-server.cnrs-mrs.fr/audic/significance.html>). Differences statistically significant at the 95% confidence level are underlined.

that malate is preferable to pyruvate as a substrate for fatty acid synthesis in non-green developing seeds such as safflower (*Carthamus tinctorius*) and castor endosperm (*Ricinus communis*) (Browse and Slack, 1985; Eastmond *et al.*, 1997).

To further explore different expression of cMDH in developing sesame and *Arabidopsis* seeds, we performed a western blot analysis with antiserum against alfalfa cMDH (Miller *et al.*, 1998). A 2- or 3-fold difference of cMDH expression was observed (Figure 4) although, since the cross-reactivity of the antibody with *Arabidopsis* and sesame cMDH is unknown, these results provide only a preliminary estimate of protein abundance. Further studies, such as stable-isotope feeding experiments (Schwender and Ohlrogge, 2002) will be required to evaluate the metabolic pathways for the generation of cofactors and carbon sources for in non-green oilseeds.

Surprisingly, only one EST encoding stearoyl-ACP desaturase and none for microsomal oleic acid desat-

urase (FAD2) were detected in current sesame ESTs even though the level of oleic acid and linoleic acid increased to 80% of total fatty acids during sesame seed development. These low levels could reflect the discrepancy between levels of transcription and translation of genes, enzyme activity, or other factors. Regardless of the explanation, the low abundance of ESTs representing these enzymes indicates that such analysis of ESTs should be considered preliminary and subject for further investigations.

Conclusion

One of the main goals of recent plant biotechnology is to manipulate metabolism in seeds for the production of improved products such as vegetable oils. Although a variety of sesame cultivars have been developed by conventional breeding methods, there is demand for genetic improvement of sesame cultivars for the generation of new varieties producing economically and

nutritionally valuable oils. In this study, 3328 sesame ESTs were generated, which has greatly expanded available sequence data for sesame and the Pedaliaceae family and in addition provides information about genes expressed during seed development and EST candidates responsible for genes involved in sesame lignan biosynthesis. We summarize the similar or different gene expression between developing seed of sesame and *Arabidopsis* ESTs as follows. First, seed storage proteins and oleosins were abundant in developing seeds of both sesame and *Arabidopsis*. In contrast, the number of ESTs encoding LTP, major latex protein, metallothionein, and allergens was substantially higher in sesame than in *Arabidopsis*, suggesting that expression of those genes is a specific characteristic of developing sesame seeds. Sesame ESTs showing no or low similarity to *Arabidopsis* proteins are of particular interest because they may correspond to genes with novel or specific roles in developing sesame seeds. Second, some specific isoforms (e.g. 12S cruciferin, oleosin, LTP, ACP, and others) in multiple gene families were preferentially expressed in sesame seeds. Third, EST candidates possibly associated with biosynthetic enzymes of sesame-specific lignans, sesamin and sesamol, were identified. Fourth, higher expression of cMDH and lower expression of plastidic pyruvate kinase in developing seeds of sesame than in *Arabidopsis* suggest that malate could be a major substrate for the generation of reducing power as well as a carbon source required for storage lipid biosynthesis in non-green seeds.

Acknowledgements

We greatly appreciate the help of Rob Halgren and Matt Larson at Michigan State University, USA, and Min Woo Nam and Ahmi Choi at Korea University with data analysis, Young-Sik Jeong at WonKwang University for construction of a website for developing sesame seed ESTs, Mike Pollard at Michigan State University for helpful discussion about sesame lignans, and Carroll P. Vance at University of Minnesota for providing antiserum against alfalfa cMDH. This work was supported in part by grants from the Plant Diversity Research Center of the 21st Century Frontier Research Program funded by the Ministry of Science and Technology of Korean Government, the BioGreen 21 Program funded by the Rural Development Administration, Republic of Korea, and the

Plant Signaling Network Research Center funded by the Korea Science and Engineering Foundation.

References

- Anterola, A.M., Jeon, J.H., Davin, L.B. and Lewis, N.G. 2002. Transcriptional control of monolignol biosynthesis in *Pinus taeda*: factors affecting monolignol ratios and carbon allocation in phenylpropanoid metabolism. *J. Biol. Chem.* 24: 18272–18280.
- Asano, T., Kunieda, N., Omura, Y., Ibe, H., Kawasaki, T., Takano, M., Sato, M., Furuhashi, H., Mujin, T., Takaiwa, F., Wu, C., Tada, Y., Satozawa, T., Sakamoto, M. and Shimada, H. 2002. Rice SPK, a calmodulin-like domain protein kinase is required for storage product accumulation during seed development: phosphorylation of sucrose synthase is a possible factor. *Plant Cell* 14: 619–628.
- Ashri, A. 1987. Sesame. In: G. Robbelen, R.K. Downey and A. Ashri (Eds.) *Oil Crops of the World: Their Breeding and Utilization*, McGraw-Hill, New York, pp. 375–387.
- Audic, S. and Claverie, J.M. 1997. The significance of digital gene expression profiles. *Genome Res.* 7: 986–995.
- Bao, X., Katz, S., Pollard, M., and Ohlrogge, J. 2002. Carbocyclic fatty acids in plants: biochemical and molecular genetic characterization of cyclopropane fatty acid synthesis of *Sterculia foetida*. *Proc. Natl Acad. Sci. USA* 14: 7172–7177.
- Browse, J. and Slack, C.R. 1985. Fatty acid synthesis in plastids from maturing safflower and linseed cotyledons. *Planta* 166: 74–80.
- Burlat, V., Kwon, M., Davin, L.B. and Lewis, N.G. 2001. Dirigent proteins and dirigent sites in lignifying tissues. *Phytochemistry* 57: 883–897.
- 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.
- Channelière, S., Riviere, S., Scalliet, G., Szecsi, J., Jullien, F., Dolle, C., Vergne, P., Dumas, C., Bendahmane, M., Huguency, P. and Cock, J.M. 2002. Analysis of gene expression in rose petals using expressed sequence tags. *FEBS Lett.* 515: 35–38.
- Chavrier, P. and Goud, B. 1999. The role of ARF and Rab GTPases in membrane transport. *Curr. Opin. Cell Biol.* 11: 466–475.
- Coonev, R.V., Custer, L.J., Okinaka, L. and Franke, A.A. 2001. Effects of dietary sesame seeds on plasma tocopherol levels. *Nutr. Cancer* 39: 66–71.
- Davin, L.B., Wang, H.B., Crowell, A.L., Bedgar, D.L., Martin, D.M., Sarkanen, S. and Lewis, N.G. 1997. Stereoselective bimolecular phenoxyl radical coupling by an auxiliary (dirigent) protein without an active center. *Science* 257: 362–367.
- Dennis, D.T. and Miernyk, J.A. 1982. Compartmentation of non-photosynthetic carbohydrate metabolism. *Annu. Rev. Plant Physiol.* 33: 27–50.
- Eastmond, P.J. and Rawsthorne, S. 2000. Coordinate changes in carbon partitioning and plastidial metabolism during the development of oilseed rape embryos. *Plant Physiol.* 122: 767–774.
- Eastmond, P.J., Dennis, D.T., and Rawsthorne, S. 1997. Evidence that a malate inorganic phosphate exchange translocator imports carbon across the leucoplast envelope for fatty acid synthesis in developing castor seed endosperm. *Plant Physiol.* 114: 851–856.

- Ewing, B., Hiller, L., Wendle, M.C. and Green, P. 1998. Base-calling of automated sequencer traces using Phred. I accuracy assessment. *Genome Res.* 8: 175–185.
- Girke, T., Todd, J., Ruuska, S., White, J., Benning, C. and Ohlrogge, J. 2000. Microarray analysis of developing *Arabidopsis* seeds. *Plant Physiol.* 124: 1570–1581.
- Harlow, E. and Lane, D. 1988. *Antibodies: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Plainview, NY.
- Jiao, Y., Davin, L.B. and Lewis, N.G. 1998. Furanofuran lignan metabolism as a function of seed maturation in *Sesamum indicum*: methylenedioxy bridge formation. *Phytochemistry* 49: 387–394.
- Kader, J.C. 1996. Lipid-transfer-proteins in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47: 627–645.
- Kato, M.J., Chu, A., Davin, L.B. and Lewis, N.G. 1998. Biosynthesis of antioxidant lignans in *Sesamum indicum* seeds. *Phytochemistry* 47: 583–591.
- Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227: 680–685.
- Marchand, P.A., Zajicek, J. and Lewis, N.G. 1997. Oxygen insertion in *Sesamum indicum* furanofuran lignans: diastereoselective syntheses of enzyme substrate analogues. *Can. J. Chem.* 75: 840–849.
- Martin, T.F.J. 1998. Phosphoinositide lipids as signaling molecules: common themes for signal transduction, cytoskeletal regulation, and membrane trafficking. *Annu. Rev. Cell. Dev. Biol.* 14: 231–264.
- McElver, J., Patton, D., Rumbaugh, M., Liu, C., Yang, L.J. and Meinke, D. 2000. The *titan5* gene of *Arabidopsis* encodes a protein related to the ADP ribosylation factor family of GTP binding proteins. *Plant Cell* 12: 1379–1392.
- Mekhedov, S., deIlarduya, O.M. and Ohlrogge, J. 2000. Toward a functional catalog of the plant genome: a survey of genes for lipid biosynthesis. *Plant Physiol.* 122: 389–402.
- Miller, S.S., Driscoll, B.T., Gregerson, R.G., Gantt, J.S. and Vance, C.P. 1998. Alfalfa malate dehydrogenase (MDH): molecular cloning and characterization of five different forms reveals a unique nodule-enhanced MDH. *Plant J.* 15: 173–184.
- Miyahara, Y., Hibasami, H., Katsuzaki, H., Imai, K. and Komiya, T. 2001. Sesamol from sesame seed inhibits proliferation by inducing apoptosis in human lymphoid leukemia Molt 4B cells. *Int. J. Mol. Med.* 7: 369–371.
- Monks, D.E., Aghoram, K., Courtney, P.D., DeWald, D.B. and Dewey, R.E. 2001. Hyperosmotic stress induces the rapid phosphorylation of a soybean phosphatidylinositol transfer protein homolog through activation of the protein kinases SPK1 and SPK2. *Plant Cell* 13: 1205–1219.
- Pastorello, E.A., Pompei, C., Pravettoni, V., Brenna, O., Farioli, L., Trambaioli, C. and Conti, A. 2001. Lipid transfer proteins and 2S albumins as allergens. *Allergy* 56 (Suppl. 67): 45–47.
- Ruuska, S.A., Girke, T., Benning, C. and Ohlrogge, J.B. 2002. Contrapuntal networks of gene expression during *Arabidopsis* seed filling. *Plant Cell* 14: 1191–1206.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. 1989. *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Plainview, NY.
- Schwender, J. and Ohlrogge, J.B. 2002. Probing *in vivo* metabolism by stable isotope labeling of storage lipids and proteins in developing *Brassica napus* embryos. *Plant Physiol.* 130: 347–361.
- Shimizu, S., Akimoto, K., Shinmen, Y., Kawashima, H., Sugano, M. and Yamada, H. 1991. Sesamin is a potent and specific inhibitor of delta-5 desaturase in polyunsaturated fatty acid biosynthesis. *Lipids* 26: 512–516.
- Tai, S.S., Wu, L.S., Chen, E.C. and Tzen, J.T. 1999. Molecular cloning of 11S globulin and 2S albumin, the two major seed storage proteins in sesame. *J. Agric. Food Chem.* 47: 4932–4938.
- Van Lijsebettens, M., Vanderhaeghen, R., De Block, M., Bauw, G., Villarroel, R. and Van Montagu, M. 1994. An S18 ribosomal protein gene copy at the *Arabidopsis* PFL locus affects plant development by its specific expression in meristems. *EMBO J.* 13: 3378–3388.
- White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de Ilarduya, O., Jaworski, J.G., Ohlrogge, J.B. and Benning, C. 2000. A new set of *Arabidopsis* expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil. *Plant Physiol.* 124: 1582–1594.
- Wirtz, K.W.A. 1991. Phospholipid transfer proteins. *Annu. Rev. Biochem.* 60: 73–99.
- Yermanos, D.M., Hemstreet, S., Saleeb, W. and Huszar, C.K. 1972. Oil content and composition of the seed in the world collection of sesame introductions. *J. Am. Oil Chem. Soc.* 49: 20–23.