Journal of Plant Biology, September 2001, 44(3) : 172-177 **REPORT**

Expressed Sequence Tags and mRNA Expression Levels of Tagged cDNAs from Watermelon Anthers and Developing Seeds

Joonyul Kim, Sung-Hoon Jun, Jinwon Lee, Hong-Gyu Kang, and Gynheung An*
Division of Molecular and Life Sciences, Pohang University of Science and Technology, Pohang 790-784, Korea

To understand the molecular events that occur during reproductive organ development and to provide genetic resources for molecular breeding, we generated 328 expressed sequence tags (ESTs) from randomly selected clones of four watermelon cDNA libraries. These libraries were prepared from young and mature anthers, as well as the seed coat and inner seed tissues. EST clones found in the young anthers and inner seed tissues showed similarity with genes related to development and signal transduction. We could deduce that, especially in the developing inner seed tissues, important morphological processes were associated exclusively with seed and embryo development. In addition, seed metabolism was tailored toward the accumulation of economically valuable storage compounds such as lipids. In the seed coat, EST clones showed similarity with genes that influence the transport or conversion of nutrients such as porin, sucrose synthase, L-asparaginase, and arginine decarboxylase. We also selected two cDNA clones from each of the four classes of ESTs for studying expression levels and patterns in the various organs. Among those eight clones, three (An88, Is124, and Sc68) were expressed preferentially in their particular organ.

Keywords: anther, developing seed, expressed sequence tags (ESTs), watermelon

Large-scale, single-pass sequencing of cDNAs from different plants has provided an extensive reservoir for the cloning of genes, evaluation of tissue-specific gene expression, use of markers for map-based cloning, and annotation of genomic sequencing. Expressed sequence tags (ESTs) have been reported for several species, including *Arabidopsis thaliana* (Newman et al., 1994; Cooke et al., 1996; White et al., 2000), rice (Sasaki et al., 1994), maize (Keith et al., 1993), *Brassica napus* (Park et al., 1993), apple (Sung et al., 1998), and radish (Moon et al., 1998). A large number of the clones analyzed by ESTs showed significant similarities to proteins identified within the same plant, or in other plants, other eukaryotes, or prokaryotes (Newman et al., 1994; Cooke et al., 1996). ESTs and their accompanying cDNAs can also provide a means for constructing inexpensive microarrays on glass slides, which can then be used to study gene expression on a genome-wide scale (DeRisi et al., 1997; Ruan et al., 1998).

A careful bioinformatic analysis to identify tissue-specific ESTs is a prerequisite for obtaining a comprehensive and representative set of cDNAs for gene expression studies by microarrays (Loftus et al., 1999). Even without that subsequent microarray analysis, a sufficiently large number of ESTs derived from a specific tissue can provide a clue toward understanding expression of specific genes in that tissue (Rafalski et al., 1998; Ewing et al., 1999; Mekhedov et al., 2000). In most cases, and within statistical limitations (Audic and Claverie, 1997), the abundance of a specific cDNA in the EST collection also is a measure of gene expression. As of January 2000, GenBank contained over 220,000 entries for ESTs from plant species. However, most of the publicly available plant ESTs have been derived from vegetative tissues; relatively few are specifically derived from reproductive organs.

In this study, we selected 328 cDNA clones that contained 0.5- to 1.5-kb inserts from cDNA libraries constructed from watermelon tissues. These included both young and mature anthers, inner seed tissues, and seed coats collected at 6 to 10 days after pollination (DAP). The average length of the deduced sequence for each clone was 150 to 250 bp. We used BLASTX to search each sequence against the non-redundant protein database of DDBJ; the top scorer was then extracted (Altschul et al., 1997). To obtain qualitative information about the ESTs, we selected the clone whose top score was >50 bits. By using this process, we found that 24 ESTs in the young anthers, 21 in mature anthers, 34 in the inner seed tissues, and 13 in the seed coat showed significant

---

*Corresponding author; fax +82-54-279-2199
e-mail genean@postech.ac.kr

Abbreviations: DAP, days after pollination; EST, expressed sequence tag.

Table 1. Watermelon ESTs showing similarities to known genes. Numbers in the columns designated "Id" refer to the percentage of amino acid sequence identity in a continuous region of a particular length (Overlap). Ea, young anthers; An, mature anthers; Is, inner seed tissues; Sc, seed coat.

EST #	Putative identification	Id (%)	Overlap (length)	Score (bit)	Source of comparison
Ea10	Calreticulin	61	71	87	*Beta vulgaris* (AJ002057)
Ea13	Ribosomal protein L29	88	45	91	*Panax ginseng* (AB042860)
Ea21	COP9 protein	68	60	95	*Arabidopsis thaliana* (Z97335)
Ea22	Chlorophyll a/b protein gene	58	50	61	Lemna gibba dad (M29334)
Ea24	Spore germination protein C2	68	35	55	*Synechocystis* sp. (D90901)
Ea26	26S proteasome	92	39	76	*Arabidopsis thaliana* (AB025633)
Ea29	Histone H2A	68	50	68	*Arabidopsis thaliana* (AF204968)
Ea32	Polyubiquitin protein	93	69	138	Saccharum hybrid cultivar (AF093505)
Ea53	Pescadillo-like protein	65	49	72	*Arabidopsis thaliana* (AL163792)
Ea65	Thaumatin-like protein	62	67	76	*Arabidopsis thaliana* (L34693)
Ea71	Trans-cinnamic acid hydroxylase	78	73	119	*Pisum sativum* (U29243)
Ea72	40S subunit ribosomal protein	76	52	95	*Oryza sativa* (D12633)
Ea74	60S ribosome protein	79	73	111	*Arabidopsis thaliana* (AB022217)
Ea76	COP9 protein	58	82	97	*Spinacia oleracea* (U51270)
Ea80	Ribosomal protein	76	63	94	*Gossypium hirsutum* (U64678)
Ea81	Sgt1 protein	76	75	115	*Oryza sativa* (AP002744)
Ea82	Ribosomal protein	71	82	115	*Gossypium hirsutum* (U64678)
Ea85	Ribosomal protein L39	86	38	75	*Zea mays* (X95458)
Ea89	Endo-1,4-beta-mannosidase	71	56	92	*Arabidopsis thaliana* (AB013389)
Ea93	Calreticulin	83	43	91	*Ricinus communis* (U74631)
Ea96	Chlorophyll a/b binding preprotein	98	61	134	*Glycine max* (X12981)
Ea100	Ubiquitin/ribosomal polyprotein	98	71	139	*Oryza sativa* (D12629)
Ea104	SAM-synthetase	88	34	68	*Cicer arietinum* (X85252)
Ea105	Casein kinase-like protein	73	69	103	*Oryza sativa* (AP003244)
An1	Cucumisin-like serine protease	57	54	69	*Arabidopsis thaliana* (AB019229)
An3	Long-chain-fatty-acid-CoA ligase	71	46	77	*Arabidopsis thaliana* (AB012247)
An6	Vacuolar ATP synthase subunit C	76	86	141	*Arabidopsis thaliana* (AF208261)
An16	Protein kinase C inhibitor	82	76	130	*Arabidopsis thaliana* (AL163972)
An24	Sec61p protein	58	70	72	*Triticum aestivum* (AF161718)
An25	Chlorophyll a/b-binding protein	76	42	70	*Cucumis sativus* (M16057)
An32	Glycine-rich RNA-binding protein	76	42	71	*Glycine max* (AF169205)
An44	Isocitrate dehydrogenase	85	62	101	*Solanum tuberosum* (X75638)
An55	Calmodulin-binding heat-shock protein	66	54	85	*Nicotiana tabacum* (S79243)
An57	Ribosomal protein S28	88	53	98	*Prunus persica* (AJ012655)
An64	Isoflavone reductase homolog 2	74	63	91	*Glycine max* (AF202184)
An66	RNA-directed RNA polymerase	76	65	119	*Petunia* x hybrida (AJ011979)
An70	Glucosyl transferase	61	54	69	*Arabidopsis thaliana* (AC073395)
An72	Phosphofructokinase beta subunit	81	55	94	*Citrus* x paradisi (AF095520)
An79	Catalase protein	82	47	73	*Cucurbita pepo* (D55647)
An80	Cinnamyl alcohol dehydrogenase	82	34	59	*Medicago sativa* (Z19573)
An88	UDP-galactose 4-epimerase	80	41	71	*Arabidopsis thaliana* (AC066689)
An92	Catalase protein.	98	85	189	*Cucurbita pepo* (D55647)
An100	Ribosomal protein S4	97	48	105	*Arabidopsis thaliana* (Z17994)
An102	Aspartate aminotransferase	81	64	106	*Glycine max* (AF034210)
An104	Nodulin-26 protein.	74	66	92	*Glycine max* (L12257)

Table 1. Continued

EST #	Putative identification	Id (%)	Overlap (length)	Score (bit)	Source of comparison
Is5	Tonoplast intrinsic protein gamma	80	60	96	*Arabidopsis thaliana* (X72581)
Is8	Ribosomal protein S15 protein.	94	59	117	*Arabidopsis thaliana* (L27461)
Is10	Protein kinase	75	45	70	*Arabidopsis thaliana* (L00670)
Is12	S-adenosyl-L-homocysteine hydrolase	77	53	83	*Catharanthus roseus* (Z26881)
Is13	Cystatin protein	53	62	68	*Lycopersicon esculentum* (AF198389)
Is19	Chlorophyll a/b binding protein	98	68	142	*Oryza sativa* (AF061577)
Is24	Putative lipid transfer protein	51	62	69	*Zinnia elegans* (U19266)
Is25	Cytochrome P450 like TBP protein	93	63	120	*Nicotiana tabacum* (D64052)
Is35	14-3-3 protein 32kDa endonuclease	91	61	110	*Cucurbita pepo* (X76086)
Is36	Multicatalytic endopeptidase	71	82	106	*Cicer arietinum* (AB025000)
Is43	Aspartic proteinase	49	51	63	*Vigna unguiculata* (U61396)
Is46	Cysteine synthase	97	41	83	*Citrullus lanatus* (D28777)
Is48	Jasmonic acid regulatory protein	80	50	91	*Arabidopsis thaliana* (AB022218)
Is52	Cysteine proteinase protein.	88	68	136	*Carica papaya* pot. (X03971)
Is54	Sex determination protein	44	103	94	*Cucumis sativus* (AF286651)
Is68	Vacuolar processing enzyme	70	112	169	*Beta vulgaris* (AJ309173)
Is69	Galactokinase	67	122	160	*Arabidopsis thaliana* (X99851)
Is70	Acyl carrier protein	57	117	112	*Casuarina glauca* (Y10994)
Is74	Omega-6 fatty acid desaturase	80	62	116	*Petroselinum crispum* (U86072)
Is75	Lleucine zipper protein	37	104	80	*Arabidopsis thaliana* (AB025632)
Is77	Cdc2-like protein kinase	84	39	66	*Arabidopsis thaliana* (AB005234)
Is80	LEDI-4 protein	59	142	161	*Lithospermum erythrorhizon* (AB030704)
Is83	P protein	82	58	111	*Pisum sativum* (X59773)
Is90	EIF3e protein.	83	107	184	*Arabidopsis thaliana* (AF255679)
Is91	Glycine-rich RNA-binding protein	73	45	74	*Glycine max* (AF169205)
Is103	Cytoplasmic aldolase	74	62	91	*Oryza sativa* (D13512)
Is107	Nodulin-26 protein.	87	40	75	*Glycine max* (L12257)
Is120	Histone H2A	71	46	64	*Arabidopsis thaliana* (AF204968)
Is124	GRAB1-like protein	79	58	112	*Arabidopsis thaliana* (AC010704)
Is132	Vacuolar proton-ATPase subunit	82	52	95	*Arabidopsis thaliana* (AC006841)
Is135	Eukaryotic initiation factor 5A3	89	59	112	*Solanum tuberosum* (AB004824)
Is141	Integral membrane protein p	95	48	92	*Phaseolus vulgaris* (AF190652)
Is151	Translation initiation factor	82	64	114	*Arabidopsis thaliana* (AB005232)
Is156	Endoxyloglucan transferase	62	80	105	*Arabidopsis thaliana* (AC084165)
Sc7	Putative zinc finger transcription factor	89	39	77	*Oryza sativa* (AP002746-9)
Sc9	Vacuolar H+-ATPase catalytic subunit	89	39	78	*Gossypium hirsutum* (L03186)
Sc11	34 kDA porin protein	90	63	123	*Solanum tuberosum* (X80386)
Sc23	S-adenosyl methionine synthetase	100	30	65	*Pinus banksiana* (U38186)
Sc32	Gibberellin 20-oxidase protein.	100	46	102	*Citrullus lanatus* (AF074709)
Sc37	Arginine decarboxylase	96	58	117	*Glycine max* (U35367)
Sc39	Cytosolic ascorbate peroxidase	85	67	122	*Fragaria* x ananassa (AF159631)
Sc43	Putative zinc finger protein	76	42	86	*Oryza sativa* (AC025783)
Sc54	Sorbitol dehydrogenase-like protein	73	57	96	*Arabidopsis thaliana* (AB015478)
Sc56	S-adenosylmethionine synthetase protein	76	56	84	*Arabidopsis thaliana* (M55077)
Sc66	O-methyltransferase protein	74	39	60	*Thalictrum tuberosum* (AF064697)
Sc68	L-asparaginase	67	43	54	*Arabidopsis thaliana* (AB012247)
Sc79	Sucrose synthase	84	50	88	*Gossypium hirsutum* (U73588)

similarities to previously identified genes (Table 1). A total of 92 clones were analyzed at the level of BLASTX searches, which was equivalent to 28% of the clones on the filter.

Clones were assigned to specific functional classes (Table 2). ESTs involved in metabolism were found in all four libraries. Although housekeeping genes, such as those coding for ribosomal proteins and ubiquitins, were found in all tissues except the seed coat, they were most abundant in young anthers. Genes involved in development and photosynthesis were also found in those latter two ESTs.

Two clones (Ea21, Ea76) in young anther ESTs had significant homology with the light signal transduction gene encoding COP9. Clones corresponding to genes encoding for enzymes that play a role in metabolism, e.g., trans-cinnamic acid hydroxylase (Ea71) and endo-1,4-beta-mannosidase (Ea89) also were found in that EST. In mature anthers, An79 and An 92 showed homology to catalase. Likewise, An80 and An88 shared homology with cinnamyl alcohol dehydrogenase and UDP-galactose 4-epimerase, respectively.

In the EST of the inner seed tissues, the most abundant classes were the clones related to development, transcription, and metabolism. Is35 (a 14-3-3 protein 32-kDa endonuclease), Is54 (the sex-determination protein), Is77 (a cdc2-like protein kinase), and Is124 (a GRAB1-like protein) were likely candidates for playing certain roles in growth regulation. In addition, Is24 and Is70 shared homology with the lipid transfer protein and acyl carrier protein, respectively, both of which affect lipid metabolism. In the seed coat, Sc32 showed high similarity with the hormone biosynthetic enzyme, GA 20-oxidase. Sc68 and Sc79 had significant homology with L-asparaginase and sucrose synthase.

For further study, we selected two clones each from the four libraries. These clones had shown significant homology with previously identified genes. For our young anther ESTs, those clones were Ea24 and Ea93, which had exhibited similarity with the spore germination protein C2 and calreticulin, respectively. We also chose An80 (cinnamyl alcohol dehydrogenase) and An88 (UDP-galactose 4-epimerase) from mature anther ESTs, Is77 (a cdc2-like protein kinase) and Is124 (a GRAB 1-like protein) from the inner seed tissue ESTs, and Sc68 (L-asparaginase) and Sc79 (sucrose synthase) from the seed coat ESTs. RNA blot analyses revealed specific expressions of An88 in mature anthers, Is124 in inner seed tissues, and Sc68 in the seed coat (Fig. 1). Sink organs in most plant species are supplied, through their seed coats, with carbon and energy sources from the phloem. In this study, both Sc68 and Sc79 appeared to preferentially regulate levels of photoassimilates.

In conclusion, EST analyses of watermelon tissues identified a number of clones that may play important roles during anther and seed development. This information will be useful for understanding the development of reproductive organs in watermelon and other plant species.

We are grateful to the Dongbu Hannong Seed Company for kindly providing the watermelon seeds.

Table 2. Distribution of cDNAs in classes of putative function. The class assignment presented is based on the description of the best match from BLASTX similarity searches to the non-redundant DDBJ protein databases.

	Young anther	Mature anther	Inner seed tissues	Seed coat	Total
Development	2	0	4	0	6
Metabolism	2	6	7	6	21
Ribosome, protein translation	6	2	2	0	10
Proteinase, ubiquitin	3	1	2	0	6
Signal transduction (kinases, calmodulin, etc.)	3	1	3	0	7
Photosynthesis	2	2	2	0	6
Transcription factor	0	0	3	2	5
Peroxidase	0	2	0	1	3
Acting on RNA	0	2	1	0	3
Vacuole	0	0	3	1	4
Secondary metabolite	1	2	1	1	5
Chaperonin	0	1	0	0	1
Membrane, transporter, receptor	0	0	1	1	2
Defense	1	0	1	0	2
Hormone biosynthetic enzyme	0	0	0	1	1
Nucleus	1	0	1	0	2
Others	3	2	3	0	8
Total	24	21	34	13	92

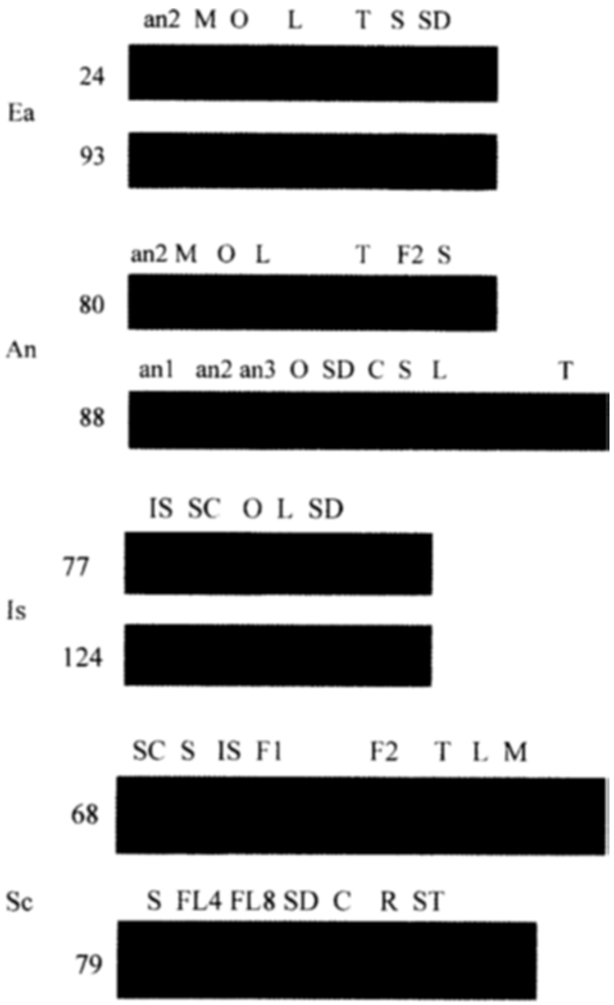

Figure 1. RNA blot analysis of tagged cDNAs that were strongly expressed in at least one reproductive organ. Total RNA was extracted from SC, seed coat (8 DAP); S, whole seed; IS, inner seed tissues (8 DAP); F1, fruits (1 DAP); F2, fruits (2-3 DAP); FL4, fleshes (4 DAP); FL8, fleshes (8 DAP); T, tendrils; L, leaves; M, male flowers; O, ovaries before pollination; an1, anthers within 1 mm of male flowers; an2, anthers within 2-3 mm of male flowers; an3, anthers within mature male flowers; SD, seedling; C, cotyledon; R, roots; ST, stem. Probes used were: Ea24 homologous to the spore germination protein C2; Ea93 homologous to calreticulin; An80 homologous to cinnamyl alcohol dehydrogenase; An88 homologous to UDP-galactose 4-epimerase; Is77 homologous to the cdc2-like protein kinase; Is124 homologous to the GRAB 1-like protein; Sc68 homologous to L-asparaginase; and Sc79 homologous to sucrose synthase.

ACKNOWLEDGMENTS

This work was funded, in part, by a grant from a special research program of the Ministry of Agriculture and Forestry of Korea. The authors thank Darongi Choi and Sun-Keun Hur for their DNA sequencing, and Chahm An and Priscilla Licht for critical reading of the manuscript.

Received October 1, 2001; accepted October 10, 2001.

LITERATURE CITED

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402

Audic S, Claverie JM (1997) The significance of digital gene expression profiles. Genome Res 10: 986-995

Cooke R, Raynal M, Laudie M, Grellet F, Delseny M, Morris P-C, Guerrier D, Giraudat J, Quigley F, Clabault G, Li Y-F, Mache R, Krivitzky M, Gy J-J, Kreis M, Lecharny A, Parmentier Y, Marbach J, Fleck J, Clement B, Philipps G, Herve C, Bardet C, Tremousaygue D, Lescure B, Lacomme C, Roby D, Jourjon M-F, Chabrier P, Charpenteau J-L, Desprez T, Amselem J, Chiapello H, Hofte H (1996) Further progress towards a catalogue of all *Arabidopsis* genes: analysis of a set of 5000 nonredundant ESTs. Plant J 9: 101-124

DeRisi JL, Iyer VR, Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278: 680-686

Ewing RM, Kahla AB, Poirot O, Lopez F, Audic S, Claverie JM (1999) Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. Genome Res 9: 950-959

Keith CS, Hoang DO, Barrett BM, Feigelman B, Nelson MC, Thai H, Baysdorfer C (1993) Partial sequencing analysis of 130 randomly selected maize cDNA clones. Plant Physiol 101: 329-332

Loftus SK, Chen Y, Gooden G, Ryan JF, Birznieks G, Hilliard M, Baxevanis AD, Bittner M, Meltzer P, Trent J, Pavan W (1999) Informatic selection of a neural crest-melanocyte cDNA set for microarray analysis. Proc Natl Acad Sci USA 96: 9277-9280

Mekhedov S, de Ilarduya OM, Ohlrogge J (2000) Toward a functional catalog of the plant genome. A survey of genes for lipid biosynthesis. Plant Physiol 122: 389-402

Moon YH, Chae S, Jung JY, An G (1998) Expressed sequence tags of radish flower buds and characterization of a CONSTANS LIKE 1 gene. Mol Cells 4: 452-458

Newman T, de Bruijn FJ, Green P, Keegstra K, Kende H, McIntosh L, Ohlrogge J, Raikhel N, Somerville S, Thomashow M, Retzel E, Somerville C (1994) Genes galore: A summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. Plant Physiol 106: 1241-1255

Park YS, Kwak JM, Kwon OY, Kim YS, Lee DS, Cho MJ, Lee HH, Nam HG (1993) Generation of expressed sequence tags of random root cDNA clones of *Brassica napus* by single-run partial sequencing. Plant Physiol

103: 359-370

Rafalski JA, Hanafey M, Miao GH, Ching A, Lee JM, Dolan M, Tingey S (1998) New experimental and computational approaches to the analysis of gene expression. Acta Biochim Pol 45: 929-934

Ruan Y, Gilmore J, Conner T (1998) Towards *Arabidopsis* genome analysis: Monitoring expression profiles of 1400 genes using cDNA microarrays. Plant J 15: 821-833

Sasaki T, Song J, Koga-Ban Y, Matsui E, Fang F, Higo H, Nagasaki H, Hori M, Miya M, Murayama-Kayano E, Takiguchi T, Takasuga A, Niki T, Ishimaru K, Lkeda H, Yamamoto Y, Mukai Y, Ohta I, Miyadera N, Havukkala I, Minobe Y (1994) Toward cataloguing all rice genes: large-scale sequencing of randomly chosen rice cDNAs from a callus cDNA library. Plant J 6: 615-624

Sung SK, Jeong DH, Nam J, Kim SH, Kim SR, An G (1998) Expressed sequence tags of fruits, peels, and carpels and analysis of mRNA expression levels of the tagged cDNAs of fruits from the Fuji apple. Mol Cells 8: 565-577

White JA, Todd J, Newman T, Focks N, Girke T, de Ilarduya OM, Jaworski JG, Ohlrogge JB, 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