



(72) THOMAS, TERRY L., US

(72) LI, ZHONGSEN, US

(71) RHONE-POULENC AGRO, FR

(51) Int.Cl.⁶ C12N 15/82, C12N 15/53, C12N 15/29, A01H 5/10, A01H 5/00

(30) 1997/04/09 (08/831,575) US

(54) **REGION REGULATRICE 5' D'OLEOSINE SERVANT A LA
MODIFICATION DE LA COMPOSITION LIPIDIQUE DE
GRAINES VEGETALES**

(54) **AN OLEOSIN 5' REGULATORY REGION FOR THE
MODIFICATION OF PLANT SEED LIPID COMPOSITION**

(57) Cette invention se rapporte à des régions régulatrices 5' d'un gène d'oléosine d'Arabidopsis. Lorsqu'elles sont liées fonctionnellement soit à la séquence de codage d'un gène hétérologue soit à une séquence complémentaire d'un gène de plantes natif, ces régions régulatrices 5' dirigent l'expression de la séquence de codage ou de la séquence complémentaire dans une graine de la plante. Ces régions régulatrices sont utiles dans des cassettes d'expression et des vecteurs d'expression pour la transformation des plantes. Cette invention présente également des procédés permettant de moduler les niveaux d'un gène hétérologue, tel qu'un gène de synthèse des acides gras ou un gène du métabolisme des lipides, en transformant une plante avec les cassettes d'expression et les vecteurs d'expression en question.

(57) The present invention is directed to 5' regulatory regions of an Arabidopsis oleosin gene. The 5' regulatory regions, when operably linked to either the coding sequence of a heterologous gene or a sequence complementary to a native plant gene, direct expression of the coding sequence or complementary sequence in a plant seed. The regulatory regions are useful in expression cassettes and expression vectors for the transformation of plants. Also provided are methods of modulating the levels of a heterologous gene such as a fatty acid synthesis or lipid metabolism gene by transforming a plant with the subject expression cassettes and expression vectors.

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/82, 15/29, 15/53, A01H 5/00, 5/10	A1	(11) International Publication Number: WO 98/45461 (43) International Publication Date: 15 October 1998 (15.10.98)
(21) International Application Number: PCT/US98/07179 (22) International Filing Date: 9 April 1998 (09.04.98) (30) Priority Data: 08/831,575 9 April 1997 (09.04.97) US (71) Applicant (for all designated States except US): RHONE-POULENC AGRO [FR/FR]; Dépt. Propriété Industrielle, 14-20, rue Pierre Baizet, F-69009 Lyon (FR). (72) Inventors; and (75) Inventors/Applicants (for US only): THOMAS, Terry, L. [US/US]; 2804 Cloister Drive, College Station, TX 77845 (US). LI, Zhongsen [CN/US]; Apartment Z-1-H, 1 Hensel, College Station, TX 77840 (US). (74) Agents: DiGIGLIO, Frank, S. et al.; Scully, Scott, Murphy & Presser, 400 Garden City Plaza, Garden City, NY 11530 (US).		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(54) Title: AN OLEOSIN 5' REGULATORY REGION FOR THE MODIFICATION OF PLANT SEED LIPID COMPOSITION (57) Abstract The present invention is directed to 5' regulatory regions of an <i>Arabidopsis</i> oleosin gene. The 5' regulatory regions, when operably linked to either the coding sequence of a heterologous gene or a sequence complementary to a native plant gene, direct expression of the coding sequence or complementary sequence in a plant seed. The regulatory regions are useful in expression cassettes and expression vectors for the transformation of plants. Also provided are methods of modulating the levels of a heterologous gene such as a fatty acid synthesis or lipid metabolism gene by transforming a plant with the subject expression cassettes and expression vectors.		

AN OLEOSIN 5' REGULATORY REGION FOR THE
MODIFICATION OF PLANT SEED LIPID COMPOSITION

BACKGROUND OF THE INVENTION

Seed oil content has traditionally been modified by plant breeding. The use of recombinant DNA technology to alter seed oil composition can accelerate this process and in some cases alter seed oils in a way that cannot be accomplished by breeding alone. The oil composition of *Brassica* has been significantly altered by modifying the expression of a number of lipid metabolism genes. Such manipulations of seed oil composition have focused on altering the proportion of endogenous component fatty acids. For example, antisense repression of the $\Delta 12$ -desaturase gene in transgenic rapeseed has resulted in an increase in oleic acid of up to 83%. Topfer et al. 1995 *Science* 268:681-686.

There have been some successful attempts at modifying the composition of seed oil in transgenic plants by introducing new genes that allow the production of a fatty acid that the host plants were not previously capable of synthesizing. Van de Loo, et al. (1995 *Proc. Natl. Acad. Sci USA* 92:6743-6747) have been able to introduce a $\Delta 12$ -hydroxylase gene into transgenic tobacco, resulting in the introduction of a novel fatty acid, ricinoleic acid, into its seed oil. The reported accumulation was modest from plants carrying constructs in which transcription of the hydroxylase gene was under the control of the cauliflower mosaic virus (CaMV) 35S promoter.

- 2 -

Similarly, tobacco plants have been engineered to produce low levels of petroselinic acid by expression of an acyl-ACP desaturase from coriander (Cahoon et al. 1992 *Proc. Natl. Acad. Sci USA* 89:11184-11188).

The long chain fatty acids (C18 and larger), have significant economic value both as nutritionally and medically important foods and as industrial commodities (Ohlrogge, J.B. 1994 *Plant Physiol.* 104:821-826). Linoleic (18:2 Δ 9,12) and α -linolenic acid (18:3 Δ 9,12,15) are essential fatty acids found in many seed oils. The levels of these fatty-acids have been manipulated in oil seed crops through breeding and biotechnology (Ohlrogge, et al. 1991 *Biochim. Biophys. Acta* 1082:1-26; Topfer et al. 1995 *Science* 268:681-686). Additionally, the production of novel fatty acids in seed oils can be of considerable use in both human health and industrial applications.

Consumption of plant oils rich in γ -linolenic acid (GLA) (18:3 Δ 6,9,12) is thought to alleviate hypercholesterolemia and other related clinical disorders which correlate with susceptibility to coronary heart disease (Brenner R.R. 1976 *Adv. Exp. Med. Biol.* 83:85-101). The therapeutic benefits of dietary GLA may result from its role as a precursor to prostaglandin synthesis (Weete, J.D. 1980 in *Lipid Biochemistry of Fungi and Other Organisms*, eds. Plenum Press, New York, pp. 59-62). Linoleic acid (18:2) (LA) is transformed into gamma linolenic acid (18:3) (GLA) by the enzyme Δ 6-desaturase.

- 3 -

Few seed oils contain GLA despite high contents of the precursor linoleic acid. This is due to the absence of Δ^6 -desaturase activity in most plants. For example, only borage (*Borago officinalis*), evening primrose (*Oenothera biennis*), and currants (*Ribes nigrum*) produce appreciable amounts of linolenic acid. Of these three species, only *Oenothera* and Borage are cultivated as a commercial source for GLA. It would be beneficial if agronomic seed oils could be engineered to produce GLA in significant quantities by introducing a heterologous Δ^6 -desaturase gene. It would also be beneficial if other expression products associated with fatty acid synthesis and lipid metabolism could be produced in plants at high enough levels so that commercial production of a particular expression product becomes feasible.

As disclosed in U.S. Patent No. 5,552,306, a cyanobacterial Δ^6 -desaturase gene has been recently isolated. Expression of this cyanobacterial gene in transgenic tobacco resulted in significant but low level GLA accumulation. (Reddy et al. 1996 *Nature Biotech.* 14:639-642). Applicant's copending U.S. Application Serial No. 08,366,779, discloses a Δ^6 -desaturase gene isolated from the plant *Borago officinalis* and its expression in tobacco under the control of the CaMV 35S promoter. Such expression resulted in significant but low level GLA and octadecatetraenoic acid (ODTA or OTA) accumulation in seeds. Thus, a need exists for a promoter which

- 4 -

functions in plants and which consistently directs high level expression of lipid metabolism genes in transgenic plant seeds.

Oleosins are abundant seed proteins associated with the phospholipid monolayer membrane of oil bodies. The first oleosin gene, L3, was cloned from maize by selecting clones whose in vitro translated products were recognized by an anti-L3 antibody (Vance et al. 1987 *J. Biol. Chem.* 262:11275-11279). Subsequently, different isoforms of oleosin genes from such different species as *Brassica*, soybean, carrot, pine, and *Arabidopsis* have been cloned (Huang, A.H.C., 1992, *Ann. Reviews Plant Phys. and Plant Mol. Biol.* 43:177-200; Kirik et al., 1996 *Plant Mol. Biol.* 31:413-417; Van Rooijen et al., 1992 *Plant Mol. Biol.* 18:1177-1179; Zou et al., *Plant Mol. Biol.* 31:429-433. Oleosin protein sequences predicted from these genes are highly conserved, especially for the central hydrophobic domain. All of these oleosins have the characteristic feature of three distinctive domains. An amphipathic domain of 40-60 amino acids is present at the N-terminus; a totally hydrophobic domain of 68-74 amino acids is located at the center; and an amphipathic α -helical domain of 33-40 amino acids is situated at the C-terminus (Huang, A.H.C. 1992).

The present invention provides 5' regulatory sequences from an oleosin gene which direct high level expression of lipid metabolism genes in transgenic plants. In accordance with the present invention,

- 5 -

chimeric constructs comprising an oleosin 5' regulatory region operably linked to coding sequence for a lipid metabolism gene such as a $\Delta 6$ -desaturase gene are provided. Transgenic plants comprising the subject chimeric constructs produce levels of GLA approaching the level found in those few plant species which naturally produce GLA such as evening primrose (*Oenothera biennis*).

SUMMARY OF THE INVENTION

The present invention is directed to 5' regulatory regions of an *Arabidopsis* oleosin gene. The 5' regulatory regions, when operably linked to either the coding sequence of a heterologous gene or sequence complementary to a native plant gene, direct expression of the heterologous gene or complementary sequence in a plant seed.

The present invention thus provides expression cassettes and expression vectors comprising an oleosin 5' regulatory region operably linked to a heterologous gene or a sequence complementary to a native plant gene.

Plant transformation vectors comprising the expression cassettes and expression vectors are also provided as are plant cells transformed by these vectors, and plants and their progeny containing the vectors.

In one embodiment of the invention, the heterologous gene or complementary gene sequence is a fatty acid synthesis gene or a lipid metabolism gene.

- 6 -

In another aspect of the present invention, a method is provided for producing a plant with increased levels of a product of a fatty acid synthesis or lipid metabolism gene.

In particular, there is provided a method for producing a plant with increased levels of a fatty acid synthesis or lipid metabolism gene by transforming a plant with the subject expression cassettes and expression vectors which comprise an oleosin 5' regulatory region and a coding sequence for a fatty acid synthesis or lipid metabolism gene.

In another aspect of the present invention, there is provided a method for cosuppressing a native fatty acid synthesis or lipid metabolism gene by transforming a plant with the subject expression cassettes and expression vectors which comprise an oleosin 5' regulatory region and a coding sequence for a fatty acid synthesis or lipid metabolism gene.

A further aspect of this invention provides a method of decreasing production of a native plant gene such as a fatty acid synthesis gene or a lipid metabolism gene by transforming a plant with an expression vector comprising a oleosin 5' regulatory region operably linked to a nucleic acid sequence complementary to a native plant gene.

Also provided are methods of modulating the levels of a heterologous gene such as a fatty acid synthesis or lipid metabolism gene by transforming a plant with the subject expression cassettes and expression vectors.

- 7 -

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 depicts the nucleotide and corresponding amino acid sequence of the borage $\Delta 6$ -desaturase gene (SEQ ID NO:1). The cytochrome b5 heme-binding motif is boxed and the putative metal binding, histidine rich motifs (HRMs) are underlined. The motifs recognized by the primers (PCR analysis) are underlined with dotted lines, i.e. tgg aaa tgg aac cat aa; and gag cat cat ttg ttt cc.

Fig. 2 is a dendrogram showing similarity of the borage $\Delta 6$ -desaturase to other membrane-bound desaturases. The amino acid sequence of the borage $\Delta 6$ -desaturase was compared to other known desaturases using Gene Works (IntelliGenetics). Numerical values correlate to relative phylogenetic distances between subgroups compared.

Fig. 3A provides a gas liquid chromatography profile of the fatty acid methyl esters (FAMES) derived from leaf tissue of a wild type tobacco 'Xanthi'.

Fig. 3B provides a gas liquid chromatography profile of the FAMES derived from leaf tissue of a tobacco plant transformed with the borage $\Delta 6$ -desaturase cDNA under transcriptional control of the CaMV 35S promoter (pAN2). Peaks corresponding to methyl linoleate (18:2), methyl γ -linolenate (18:3 γ), methyl α -linolenate (18:3 α), and methyl octadecatetraenoate (18:4) are indicated.

- 8 -

Fig. 4 is the nucleotide sequence and corresponding amino acid sequence of the oleosin AtS21 cDNA (SEQ ID NO:3).

Fig. 5 is an acidic-base map of the predicted AtS21 protein generated by DNA Strider 1.2.

Fig. 6 is a Kyte-Doolittle plot of the predicted AtS21 protein generated by DNA Strider 1.2.

Fig. 7 is a sequence alignment of oleosins isolated from *Arabidopsis*. Oleosin sequences published or deposited in EMBL, BCM, NCBI databases were aligned to each other using GeneWorks® 2.3. Identical residues are boxed with rectangles. The seven sequences fall into three groups. The first group includes AtS21 (SEQ ID NO:5), X91918 (SEQ ID NO:6) and Z29859 (SEQ ID NO:7). The second group includes X62352 (SEQ ID NO:8) and Atol3 (SEQ ID NO:9). The third group includes X91956 (SEQ ID NO:10) and L40954 (SEQ ID NO:11). Differences in amino acid residues within the same group are indicated by shadows. Ato2/Z54164 is identical to AtS21. Atol3 sequence (Accession No. Z541654 in EMBL database) is actually not disclosed in the EMBL database. The Z54165 Accession number designates the same sequence as Z54164 which is Ato12.

Fig. 8A is a Northern analysis of the AtS21 gene. An RNA gel blot containing ten micrograms of total RNA extracted from *Arabidopsis* flowers (F), leaves (L), roots (R), developing seeds (Se), and developing silique coats (Si) was hybridized with a probe made from the full-length AtS21 cDNA.

- 9 -

Fig. 8B is a Southern analysis of the AtS21 gene. A DNA gel blot containing ten micrograms of genomic DNA digested with BamHI (B), EcoRI (E), HindIII (H), SacI (S), and XbaI (X) was hybridized with a probe made from the full length AtS21 cDNA.

Fig. 9 is the nucleotide sequence of the SacI fragment of AtS21 genomic DNA (SEQ ID NO:12). The promoter and intron sequences are in uppercase. The fragments corresponding to AtS21 cDNA sequence are in lower case. The first ATG codon and a putative TATA box are shadowed. The sequence complementary to 21P primer for PCR amplification is boxed. A putative abscisic acid response element (ABRE) and two 14 bp repeats are underlined.

Fig. 10 is a map of AtS21 promoter/GUS construct (pAN5).

Fig. 11A depicts AtS21/GUS gene expression in *Arabidopsis* bolt and leaves.

Fig. 11B depicts AtS21 GUS gene expression in *Arabidopsis* siliques.

Fig. 11C depicts AtS21 GUS gene expression in *Arabidopsis* developing seeds.

Figs. 11D through 11J depict AtS21 GUS gene expression in *Arabidopsis* developing embryos.

Fig. 11K depicts AtS21/GUS gene expression in *Arabidopsis* root and root hairs of a young seedling.

Fig. 11L depicts AtS21/GUS gene expression in *Arabidopsis* cotyledons and the shoot apex of a five day seedling.

-10-

Figs. 11M and 11N depict AtS21/GUS gene expression in *Arabidopsis* cotyledons and the shoot apex of 5-15 day seedlings.

Fig. 12A depicts AtS21/GUS gene expression in tobacco embryos and endosperm.

Fig. 12B depicts AtS21/GUS gene expression in germinating tobacco seeds.

Fig. 12C depicts AtS21/GUS gene expression in a 5 day old tobacco seedling.

Fig. 12D depicts AtS21/GUS gene expression in 5-15 day old tobacco seedlings.

Fig. 13A is a Northern analysis showing AtS21 mRNA levels in developing wild-type *Arabidopsis* seedlings. Lane 1 was loaded with RNA from developing seeds, lane 2 was loaded with RNA from seeds imbibed for 24-48 hours, lane 3: 3 day seedlings; lane 4: 4 day seedlings; lane 5: 5 day seedlings; lane 6: 6 day seedlings; lane 7: 9 day seedlings; lane 8: 12 day seedlings. Probe was labeled AtS21 cDNA. Exposure was for one hour at -80°C.

Fig. 13B is the same blot as Fig. 13A only exposure was for 24 hours at -80°C.

Fig. 13C is the same blot depicted in Figs. 13A and 13B after stripping and hybridization with an *Arabidopsis* tubulin gene probe. The small band in each of lanes 1 and 2 is the remnant of the previous AtS21 probe. Exposure was for 48 hours at -80°C.

Fig. 14 is a graph comparing GUS activities expressed by the AtS21 and 35S promoters. GUS activities expressed by the AtS21 promoter in

-11-

developing *Arabidopsis* seeds and leaf are plotted side by side with those expressed by the 35S promoter. The GUS activities expressed by the AtS21 promoter in tobacco dry seed and leaf are plotted on the right side of the figure. GUS activity in tobacco leaf is so low that no column appears. "G-H" denotes globular to heart stage; "H-T" denotes heart to torpedo stage; "T-C" denotes torpedo to cotyledon stage; "Early C" denotes early cotyledon; "Late C" denotes late cotyledon. The standard deviations are listed in Table 2.

Fig. 15A is an RNA gel blot analysis carried out on 5 µg samples of RNA isolated from borage leaf, root, and 12 dpp embryo tissue, using labeled borage $\Delta 6$ -desaturase cDNA as a hybridization probe.

Fig. 15B depicts a graph corresponding to the Northern analysis results for the experiment shown in Fig. 15A.

Fig. 16A is a graph showing relative legumin RNA accumulation in developing borage embryos based on results of Northern blot.

Fig. 16B is a graph showing relative oleosin RNA accumulation in developing borage embryos based on results of Northern blot.

Fig. 16C is a graph showing relative $\Delta 6$ -desaturase RNA accumulation in developing borage embryos based on results of Northern blot.

Fig. 17 is a PCR analysis showing the presence of the borage delta 6-desaturase gene in transformed plants of oilseed rape. Lanes 1, 3 and 4

-12-

were loaded with PCR reactions performed with DNA from plants transformed with the borage delta 6-desaturase gene linked to the oleosin 5' regulatory region; lane 2: DNA from plant transformed with the borage delta 6-desaturase gene linked to the albumin 5' regulatory region; lanes 5 and 6: DNA from non-transformed plants; lane 7: molecular weight marker (1 kb ladder, Gibco BRL); lane 8: PCR without added template DNA; lane 9: control with DNA from *Agrobacterium tumefaciens* EHA 105 containing the plasmid pAN3 (i.e. the borage delta6-desaturase gene linked to the oleosin 5' regulatory region).

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides isolated nucleic acids encoding 5' regulatory regions from an *Arabidopsis* oleosin gene. In accordance with the present invention, the subject 5' regulatory regions, when operably linked to either a coding sequence of a heterologous gene or a sequence complementary to a native plant gene, direct expression of the coding sequence or complementary sequence in a plant seed. The oleosin 5' regulatory regions of the present invention are useful in the construction of an expression cassette which comprises in the 5' to 3' direction, a subject oleosin 5' regulatory region, a heterologous gene or sequence complementary to a native plant gene under control of the regulatory region and a 3' termination sequence. Such an expression cassette can be incorporated into a variety

-13-

of autonomously replicating vectors in order to construct an expression vector.

It has been surprisingly found that plants transformed with the expression vectors of the present invention produce levels of GLA approaching the level found in those few plant species which naturally produce GLA such as evening primrose (*Oenothera biennis*).

As used herein, the term "cassette" refers to a nucleotide sequence capable of expressing a particular gene if said gene is inserted so as to be operably linked to one or more regulatory regions present in the nucleotide sequence. Thus, for example, the expression cassette may comprise a heterologous coding sequence which is desired to be expressed in a plant seed. The expression cassettes and expression vectors of the present invention are therefore useful for directing seed-specific expression of any number of heterologous genes. The term "seed-specific expression" as used herein, refers to expression in various portions of a plant seed such as the endosperm and embryo.

An isolated nucleic acid encoding a 5' regulatory region from an oleosin gene can be provided as follows. Oleosin recombinant genomic clones are isolated by screening a plant genomic DNA library with a cDNA (or a portion thereof) representing oleosin mRNA. A number of different oleosin cDNAs have been isolated. The methods used to isolate such cDNAs as well as the nucleotide and corresponding amino acid

- 14 -

sequences have been published in Kirik et al. 1986 *Plant Mol. Biol.* 31:413-417; Zou et al. *Plant Mol. Biol.* 31:429-433; Van Rooijen et al. 1992 *Plant Mol. Biol.* 18:1177-1179.

Virtual subtraction screening of a tissue specific library using a random primed polymerase chain (RP-PCR) cDNA probe is another method of obtaining an oleosin cDNA useful for screening a plant genomic DNA library. Virtual subtraction screening refers to a method where a cDNA library is constructed from a target tissue and displayed at a low density so that individual cDNA clones can be easily separated. These cDNA clones are subtractively screened with driver quantities (i.e., concentrations of DNA to kinetically drive the hybridization reaction) of cDNA probes made from tissue or tissues other than the target tissue (i.e. driver tissue). The hybridized plaques represent genes that are expressed in both the target and the driver tissues; the unhybridized plaques represent genes that may be target tissue-specific or low abundant genes that can not be detected by the driver cDNA probe. The unhybridized cDNAs are selected as putative target tissue-specific genes and further analyzed by one-pass sequencing and Northern hybridization.

Random primed PCR (RP-PCR) involves synthesis of large quantities of cDNA probes from a trace amount of cDNA template. The method combines the amplification power of PCR with the representation

-15-

of random priming to simultaneously amplify and label double-stranded cDNA in a single tube reaction.

Methods considered useful in obtaining oleosin genomic recombinant DNA are provided in Sambrook et al. 1989, in *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, NY, for example, or any of the myriad of laboratory manuals on recombinant DNA technology that are widely available. To determine nucleotide sequences, a multitude of techniques are available and known to the ordinarily skilled artisan. For example, restriction fragments containing an oleosin regulatory region can be subcloned into the polylinker site of a sequencing vector such as pBluescript (Stratagene). These pBluescript subclones can then be sequenced by the double-stranded dideoxy method (Chen and Seeburg, 1985, *DNA* 4:165).

In a preferred embodiment, the oleosin regulatory region comprises nucleotides 1-1267 of Fig. 9 (SEQ ID NO:12). Modifications to the oleosin regulatory region as set forth in SEQ ID NO:12 which maintain the characteristic property of directing seed-specific expression, are within the scope of the present invention. Such modifications include insertions, deletions and substitutions of one or more nucleotides.

The 5' regulatory region of the present invention can be derived from restriction endonuclease or exonuclease digestion of an oleosin genomic clone. Thus, for example, the known nucleotide or amino acid

-16-

sequence of the coding region of an isolated oleosin gene (e.g. Fig. 7) is aligned to the nucleic acid or deduced amino acid sequence of an isolated oleosin genomic clone and 5' flanking sequence (i.e., sequence upstream from the translational start codon of the coding region) of the isolated oleosin genomic clone located.

The oleosin 5' regulatory region as set forth in SEQ ID NO:12 (nucleotides 1-1267 of Fig. 9) may be generated from a genomic clone having either or both excess 5' flanking sequence or coding sequence by exonuclease III-mediated deletion. This is accomplished by digesting appropriately prepared DNA with exonuclease III (exoIII) and removing aliquots at increasing intervals of time during the digestion. The resulting successively smaller fragments of DNA may be sequenced to determine the exact endpoint of the deletions. There are several commercially available systems which use exonuclease III (exoIII) to create such a deletion series, e.g. Promega Biotech, "Erase-A-Base" system. Alternatively, PCR primers can be defined to allow direct amplification of the subject 5' regulatory regions.

Using the same methodologies, the ordinarily skilled artisan can generate one or more deletion fragments of nucleotides 1-1267 as set forth in SEQ ID NO:12. Any and all deletion fragments which comprise a contiguous portion of nucleotides set forth in SEQ ID NO:12 and which retain the capacity to

-17-

direct seed-specific expression are contemplated by the present invention.

The identification of oleosin 5' regulatory sequences which direct seed-specific expression comprising nucleotides 1-1267 of SEQ ID NO:12 and modifications or deletion fragments thereof, can be accomplished by transcriptional fusions of specific sequences with the coding sequences of a heterologous gene, transfer of the chimeric gene into an appropriate host, and detection of the expression of the heterologous gene. The assay used to detect expression depends upon the nature of the heterologous sequence. For example, reporter genes, exemplified by chloramphenicol acetyl transferase and β -glucuronidase (GUS), are commonly used to assess transcriptional and translational competence of chimeric constructions. Standard assays are available to sensitively detect the reporter enzyme in a transgenic organism. The β -glucuronidase (GUS) gene is useful as a reporter of promoter activity in transgenic plants because of the high stability of the enzyme in plant cells, the lack of intrinsic β -glucuronidase activity in higher plants and availability of a quantitative fluorimetric assay and a histochemical localization technique. Jefferson et al. (1987 *EMBO J* 6:3901) have established standard procedures for biochemical and histochemical detection of GUS activity in plant tissues. Biochemical assays are performed by mixing plant tissue lysates with 4-methylumbelliferyl- β -D-glucuronide, a fluorimetric substrate for GUS, incubating one hour at 37°C, and

-18-

then measuring the fluorescence of the resulting 4-methyl-umbelliferone. Histochemical localization for GUS activity is determined by incubating plant tissue samples in 5-bromo-4-chloro-3-indolyl-glucuronide (X-Gluc) for about 18 hours at 37°C and observing the staining pattern of X-Gluc. The construction of such chimeric genes allows definition of specific regulatory sequences and demonstrates that these sequences can direct expression of heterologous genes in a seed-specific manner.

Another aspect of the invention is directed to expression cassettes and expression vectors (also termed herein "chimeric genes") comprising a 5' regulatory region from an oleosin gene which directs seed specific expression operably linked to the coding sequence of a heterologous gene such that the regulatory element is capable of controlling expression of the product encoded by the heterologous gene. The heterologous gene can be any gene other than oleosin. If necessary, additional regulatory elements or parts of these elements sufficient to cause expression resulting in production of an effective amount of the polypeptide encoded by the heterologous gene are included in the chimeric constructs.

Accordingly, the present invention provides chimeric genes comprising sequences of the oleosin 5' regulatory region that confer seed-specific expression which are operably linked to a sequence encoding a heterologous gene such as a lipid metabolism enzyme.

- 19 -

Examples of lipid metabolism genes useful for practicing the present invention include lipid desaturases such as $\Delta 6$ -desaturases, $\Delta 12$ -desaturases, $\Delta 15$ -desaturases and other related desaturases such as stearyl-ACP desaturases, acyl carrier proteins (ACPs), thioesterases, acetyl transacylases, acetyl-coA carboxylases, ketoacyl-synthases, malonyl transacylases, and elongases. Such lipid metabolism genes have been isolated and characterized from a number of different bacteria and plant species. Their nucleotide coding sequences as well as methods of isolating such coding sequences are disclosed in the published literature and are widely available to those of skill in the art.

In particular, the $\Delta 6$ -desaturase genes disclosed in U.S. Patent No. 5,552,306 and applicants' copending U.S. Application Serial No. 08/366,779 filed December 30, 1994 and incorporated herein by reference, are contemplated as lipid metabolism genes particularly useful in the practice of the present invention.

The chimeric genes of the present invention are constructed by ligating a 5' regulatory region of a oleosin genomic DNA to the coding sequence of a heterologous gene. The juxtaposition of these sequences can be accomplished in a variety of ways. In a preferred embodiment the order of the sequences, from 5' to 3', is an oleosin 5' regulatory region (including a promoter), a coding sequence, and a

- 20 -

termination sequence which includes a polyadenylation site.

Standard techniques for construction of such chimeric genes are well known to those of ordinary skill in the art and can be found in references such as Sambrook et al. (1989). A variety of strategies are available for ligating fragments of DNA, the choice of which depends on the nature of the termini of the DNA fragments. One of ordinary skill in the art recognizes that in order for the heterologous gene to be expressed, the construction requires promoter elements and signals for efficient polyadenylation of the transcript. Accordingly, the oleosin 5' regulatory region that contains the consensus promoter sequence known as the TATA box can be ligated directly to a promoterless heterologous coding sequence.

The restriction or deletion fragments that contain the oleosin TATA box are ligated in a forward orientation to a promoterless heterologous gene such as the coding sequence of β -glucuronidase (GUS). The skilled artisan will recognize that the subject oleosin 5' regulatory regions can be provided by other means, for example chemical or enzymatic synthesis. The 3' end of a heterologous coding sequence is optionally ligated to a termination sequence comprising a polyadenylation site, exemplified by, but not limited to, the nopaline synthase polyadenylation site, or the octopine T-DNA gene 7 polyadenylation site. Alternatively, the polyadenylation site can be provided by the heterologous gene.

The present invention also provides methods of increasing levels of heterologous genes in plant seeds. In accordance with such methods, the subject expression cassettes and expression vectors are introduced into a plant in order to effect expression of a heterologous gene. For example, a method of producing a plant with increased levels of a product of a fatty acid synthesis or lipid metabolism gene is provided by transforming a plant cell with an expression vector comprising an oleosin 5' regulatory region operably linked to a fatty acid synthesis or lipid metabolism gene and regenerating a plant with increased levels of the product of said fatty acid synthesis or lipid metabolism gene.

Another aspect of the present invention provides methods of reducing levels of a product of a gene which is native to a plant which comprises transforming a plant cell with an expression vector comprising a subject oleosin regulatory region operably linked to a nucleic acid sequence which is complementary to the native plant gene. In this manner, levels of endogenous product of the native plant gene are reduced through the mechanism known as antisense regulation. Thus, for example, levels of a product of a fatty acid synthesis gene or lipid metabolism gene are reduced by transforming a plant with an expression vector comprising a subject oleosin 5' regulatory region operably linked to a nucleic acid sequence which is complementary to a nucleic acid

- 22 -

sequence coding for a native fatty acid synthesis or lipid metabolism gene.

The present invention also provides a method of cosuppressing a gene which is native to a plant which comprises transforming a plant cell with an expression vector comprising a subject oleosin 5' regulatory region operably linked to a nucleic acid sequence coding for the native plant gene. In this manner, levels of endogenous product of the native plant gene are reduced through the mechanism known as cosuppression. Thus, for example, levels of a product of a fatty acid synthesis gene or lipid metabolism gene are reduced by transforming a plant with an expression vector comprising a subject oleosin 5' regulatory region operably linked to a nucleic acid sequence coding for a native fatty acid synthesis or lipid metabolism gene native to the plant. Although the exact mechanism of cosuppression is not completely understood, one skilled in the art is familiar with published works reporting the experimental conditions and results associated with cosuppression (Napoli et al. 1990 *The Plant Cell* 2:270-289; Van der Krol 1990 *The Plant Cell* 2:291-299).

To provide regulated expression of the heterologous or native genes, plants are transformed with the chimeric gene constructions of the invention. Methods of gene transfer are well known in the art. The chimeric genes can be introduced into plants by leaf disk transformation-regeneration procedure as described by Horsch et al. 1985 *Science* 227:1229.

- 23 -

Other methods of transformation such as protoplast culture (Horsch et al. 1984 *Science* 223:496, DeBlock et al. 1984 *EMBO J.* 2:2143, Barton et al. 1983, *Cell* 32:1033) can also be used and are within the scope of this invention. In a preferred embodiment, plants are transformed with *Agrobacterium*-derived vectors such as those described in Klett et al. (1987) *Annu. Rev. Plant Physiol.* 38:467. Other well-known methods are available to insert the chimeric genes of the present invention into plant cells. Such alternative methods include biolistic approaches (Klein et al. 1987 *Nature* 327:70), electroporation, chemically-induced DNA uptake, and use of viruses or pollen as vectors.

When necessary for the transformation method, the chimeric genes of the present invention can be inserted into a plant transformation vector, e.g. the binary vector described by Bevan, M. 1984 *Nucleic Acids Res.* 12:8711-8721. Plant transformation vectors can be derived by modifying the natural gene transfer system of *Agrobacterium tumefaciens*. The natural system comprises large Ti (tumor-inducing) plasmids containing a large segment, known as T-DNA, which is transferred to transformed plants. Another segment of the Ti plasmid, the *vir* region, is responsible for T-DNA transfer. The T-DNA region is bordered by terminal repeats. In the modified binary vectors, the tumor inducing genes have been deleted and the functions of the *vir* region are utilized to transfer foreign DNA bordered by the T-DNA border sequences. The T-region also contains a selectable

- 24 -

marker for antibiotic resistance, and a multiple cloning site for inserting sequences for transfer. Such engineered strains are known as "disarmed" *A. tumefaciens* strains, and allow the efficient transfer of sequences bordered by the T-region into the nuclear genome of plants.

Surface-sterilized leaf disks and other susceptible tissues are inoculated with the "disarmed" foreign DNA-containing *A. tumefaciens*, cultured for a number of days, and then transferred to antibiotic-containing medium. Transformed shoots are then selected after rooting in medium containing the appropriate antibiotic, and transferred to soil. Transgenic plants are pollinated and seeds from these plants are collected and grown on antibiotic medium.

Expression of a heterologous or reporter gene in developing seeds, young seedlings and mature plants can be monitored by immunological, histochemical or activity assays. As discussed herein, the choice of an assay for expression of the chimeric gene depends upon the nature of the heterologous coding region. For example, Northern analysis can be used to assess transcription if appropriate nucleotide probes are available. If antibodies to the polypeptide encoded by the heterologous gene are available, Western analysis and immunohistochemical localization can be used to assess the production and localization of the polypeptide. Depending upon the heterologous gene, appropriate biochemical assays can be used. For example,

- 25 -

acetyltransferases are detected by measuring acetylation of a standard substrate. The expression of a lipid desaturase gene can be assayed by analysis of fatty acid methyl esters (FAMES).

Another aspect of the present invention provides transgenic plants or progeny of these plants containing the chimeric genes of the invention. Both monocotyledonous and dicotyledonous plants are contemplated. Plant cells are transformed with the chimeric genes by any of the plant transformation methods described above. The transformed plant cell, usually in the form of a callus culture, leaf disk, explant or whole plant (via the vacuum infiltration method of Bechtold et al. 1993 *C.R. Acad. Sci. Paris*, 316:1194-1199) is regenerated into a complete transgenic plant by methods well-known to one of ordinary skill in the art (e.g. Horsch et al. 1985 *Science* 227:1129). In a preferred embodiment, the transgenic plant is sunflower, cotton, oil seed rape, maize, tobacco, *Arabidopsis*, peanut or soybean. Since progeny of transformed plants inherit the chimeric genes, seeds or cuttings from transformed plants are used to maintain the transgenic line.

The following examples further illustrate the invention.

- 26 -

EXAMPLE 1**Isolation of Membrane-Bound Polysomal
RNA and Construction of Borage cDNA Library**

Membrane-bound polysomes were isolated from borage seeds 12 days post pollination (12 DPP) using the protocol established for peas by Larkins and Davies (1975 *Plant Phys.* 55: 749-756). RNA was extracted from the polysomes as described by Mechler (1987 *Methods in Enzymology* 152: 241-248, Academic Press). Poly-A⁺ RNA was isolated from the membrane bound polysomal RNA using Oligotex-dT[™] beads (Qiagen).

Corresponding cDNA was made using Stratagene's ZAP cDNA synthesis kit. The cDNA library was constructed in the lambda ZAP II vector (Stratagene) using the lambda ZAP II kit. The primary library was packaged with Gigapack II Gold packaging extract (Stratagene).

- 27 -

EXAMPLE 2Isolation of a Δ -6 Desaturase cDNA from BorageHybridization protocol

The amplified borage cDNA library was plated at low density (500 pfu on 150 mm petri dishes). Highly prevalent seed storage protein cDNAs were reduced (subtracted from the total cDNAs) by screening with the corresponding cDNAs.

Hybridization probes for screening the borage cDNA library were generated by using random primed DNA synthesis as described by Ausubel et al (1994 Current Protocols in Molecular Biology, Wiley Interscience, N.Y.) and corresponded to previously identified abundantly expressed seed storage protein cDNAs. Unincorporated nucleotides were removed by use of a G-50 spin column (Boehringer Mannheim). Probe was denatured for hybridization by boiling in a water bath for 5 minutes, then quickly cooled on ice. Nitrocellulose filters carrying fixed recombinant bacteriophage were prehybridized at 60°C for 2-4 hours in hybridization solution [4X SET (600 mM NaCl, 80 mM Tris-HCl, 4 mM Na₂EDTA; pH 7.8), 5X Denhardt's reagent (0.1% bovine serum albumin, 0.1% Ficoll, and 0.1% polyvinylpyrrolidone), 100 µg/ml denatured salmon sperm DNA, 50 µg/ml polyadenine and 10 µg/ml polycytidine]. This was replaced with fresh hybridization solution to which denatured radioactive probe (2 ng/ml hybridization solution) was added. The filters were incubated at 60°C with agitation overnight. Filters

- 28 -

were washed sequentially in 4X, 2X, and 1X SET (150 mM NaCl, 20 mM Tris-HCl, 1 mM Na₂EDTA; pH7.8) for 15 minutes each at 60°C. Filters were air dried and then exposed to X-ray film for 24 hours with intensifying screens at -80°C.

Non-hybridizing plaques were excised using Stratagene's excision protocol and reagents. Resulting bacterial colonies were used to inoculate liquid cultures and were either sequenced manually or by an ABI automated sequencer.

Random Sequencing of cDNAs from a Borage Seed 12 (DPP) Membrane-Bound Polysomal Library

Each cDNA corresponding to a non-hybridizing plaque was sequenced once and a sequence tag generated from 200-300 base pairs. All sequencing was performed by cycle sequencing (Epicentre). Over 300 expressed sequence tags (ESTs) were generated. Each sequence tag was compared to the GenBank database using the BLAST algorithm (Altschul et al. 1990 *J. Mol. Biol.* 215:403-410). A number of lipid metabolism genes, including the $\Delta 6$ -desaturase were identified.

Database searches with the cDNA clone designated mbp-65 using BLASTX with the GenBank database resulted in a significant match to the previously isolated *Synechocystis* $\Delta 6$ -desaturase. It was determined however, that mbp-65 was not a full length cDNA. A full length cDNA was isolated using mbp-65 to screen the borage membrane-bound polysomal library. The resultant clone was designated pAN1 and the cDNA insert of pAN1 was sequenced by the cycle

- 29 -

sequencing method. The amino acid sequence deduced from the open reading frame (Fig. 1, SEQ ID NO:1) was compared to other known desaturases using Geneworks (IntelligGenetics) protein alignment program. This alignment indicated that the cDNA insert of pAN1 was the borage Δ^6 -desaturase gene.

The resulting dendrogram (Figure 2) shows that Δ^{15} -desaturases and Δ^{12} -desaturases comprise two groups. The newly isolated borage sequence and the previously isolated *Synechocystis* Δ^6 -desaturase (U.S. Patent No. 5,552,306) formed a third distinct group. A comparison of amino acid motifs common to desaturases and thought to be involved catalytically in metal binding illustrates the overall similarity of the protein encoded by the borage gene to desaturases in general and the *Synechocystis* Δ^6 -desaturase in particular (Table 1). At the same time, comparison of the motifs in Table 1 indicates definite differences between this protein and other plant desaturases. Furthermore, the borage sequence is also distinguished from known plant membrane associated fatty acid desaturases by the presence of a heme binding motif conserved in cytochrome b_5 proteins (Schmidt et al. 1994 *Plant Mol. Biol.* 26:631-642) (Figure 1). Thus, while these results clearly suggested that the isolated cDNA was a borage Δ^6 -desaturase gene, further confirmation was necessary. To confirm the identity of the borage Δ^6 -desaturase cDNA, the cDNA insert from pAN1 was cloned into an expression cassette for stable expression. The vector pBI121 (Jefferson et al. 1987

- 30 -

EMBO J. 6:3901-3907) was prepared for ligation by digestion with BamHI and EcoICR I (an isoschizomer of SacI which leaves blunt ends; available from Promega) which excises the GUS coding region leaving the 35S promoter and NOS terminator intact. The borage Δ^6 -desaturase cDNA was excised from the recombinant plasmid (pAN1) by digestion with BamHI and XhoI. The XhoI end was made blunt by performing a fill-in reaction catalyzed by the Klenow fragment of DNA polymerase I. This fragment was then cloned into the BamHI/EcoICR I sites of pBI121.1, resulting in the plasmid pAN2.

TABLE 1

COMPARISON OF COMMON AMINO ACID MOTIFS IN MEMBRANE-BOUND DESATURASES

<u>Desaturase</u>	<u>Lipid Box</u>	<u>Metal Box 1</u>	<u>Metal Box 2</u>
Borage Δ^6	WIGHDAGH (SEQ. ID. NO:15)	HNAHH (SEQ. ID. NO:21)	FQIEHH (SEQ. ID. NO:29)
Synechocystis Δ^6	NVGHDANH (SEQ. ID. NO:16)	HNYLHH (SEQ. ID. NO:22)	HQVTHH (SEQ. ID. NO:30)
Arab. chloroplast Δ^{15}	VLGHDCGH (SEQ. ID. NO:17)	HRTHH (SEQ. ID. NO:23)	HVIHH (SEQ. ID. NO:31)
Rice Δ^{15}	VLGHDCGH (SEQ. ID. NO:17)	HRTHH (SEQ. ID. NO:23)	HVIHH (SEQ. ID. NO:31)
Glycine chloroplast Δ^{15}	VLGHDCGH (SEQ. ID. NO:17)	HRTHH (SEQ. ID. NO:23)	HVIHH (SEQ. ID. NO:31)
Arab. fad3 (Δ^{15})	VLGHDCGH (SEQ. ID. NO:17)	HRTHH (SEQ. ID. NO:23)	HVIHH (SEQ. ID. NO:31)
Brassica fad 3 (Δ^{15})	VLGHDCGH (SEQ. ID. NO:17)	HRTHH (SEQ. ID. NO:23)	HVIHH (SEQ. ID. NO:31)
Borage Δ^{12} (P1-81)*	VIAHECGH (SEQ. ID. NO:18)	HRRHH (SEQ. ID. NO:24)	HVAHH (SEQ. ID. NO:32)
Arab. fad2 (Δ^{12})	VIAHECGH (SEQ. ID. NO:18)	HRRHH (SEQ. ID. NO:24)	HVAHH (SEQ. ID. NO:32)
Arab. chloroplast Δ^{12}	VIGHDCAH (SEQ. ID. NO:19)	HDRHH (SEQ. ID. NO:25)	HIPHH (SEQ. ID. NO:33)
Glycine plastid Δ^{12}	VIGHDCAH (SEQ. ID. NO:19)	HDRHH (SEQ. ID. NO:25)	HIPHH (SEQ. ID. NO:33)
Spinach plastidial n-6	VIGHDCAH (SEQ. ID. NO:19)	HDQHH (SEQ. ID. NO:26)	HIPHH (SEQ. ID. NO:33)
Synechocystis Δ^{12}	VVGHDCCGH (SEQ. ID. NO:20)	HDHHH (SEQ. ID. NO:27)	HIPHH (SEQ. ID. NO:33)
Anabaena Δ^{12}	VLGHDCGH (SEQ. ID. NO:17)	HNHHH (SEQ. ID. NO:28)	HVPHH (SEQ. ID. NO:34)

*p1-81 is a full length cDNA which was identified by EST analysis and shows high similarity to the Arbidopsis Δ^{12} desaturase (fad2)

- 32 -

EXAMPLE 3

Production of Transgenic
Plants and Preparation and
Analysis of Fatty Acid Methyl Esters (FAMES)

The expression plasmid, pAN2 was used to transform tobacco (*Nicotiana tabacum* cv. *xanthi*) via *Agrobacterium tumefaciens* according to standard procedures (Horsch, et al. 1985 *Science* 227:1229-1231; Bogue et al. 1990 *Mol. Gen. Genet.* 221:49-57) except that the initial transformants were selected on 100 μ g/ml kanamycin.

Tissue from transgenic plants was frozen in liquid nitrogen and lyophilized overnight. FAMES were prepared as described by Dahmer, et al. (1989) *J. Amer. Oil. Chem. Soc.* 66: 543-548. In some cases, the solvent was evaporated again, and the FAMES were resuspended in ethyl acetate and extracted once with deionized water to remove any water soluble contaminants. FAMES were analyzed using a Tracor-560 gas liquid chromatograph as previously described (Reddy et al. 1996 *Nature Biotech.* 14:639-642).

As shown in Figure. 3, transgenic tobacco leaves containing the borage cDNA produced both GLA and octadecatetraenoic acid (OTA) (18:4 Δ 6,9,12,15). These results thus demonstrate that the isolated cDNA encodes a borage Δ 6-desaturase.

EXAMPLE 4**Expression of $\Delta 6$ -desaturase in Borage**

The native expression of $\Delta 6$ -desaturase was examined by Northern Analysis of RNA derived from borage tissues. RNA was isolated from developing borage embryos following the method of Chang et al. 1993 *Plant Mol. Biol. Rep.* 11:113-116. RNA was electrophoretically separated on formaldehyde-agarose gels, blotted to nylon membranes by capillary transfer, and immobilized by baking at 80°C for 30 minutes following standard protocols (Brown T., 1996 in *Current Protocols in Molecular Biology*, eds. Auselbel, et al. [Greene Publishing and Wiley-Interscience, New York] pp. 4.9.1-4.9.14.). The filters were preincubated at 42°C in a solution containing 50% deionized formamide, 5X Denhardt's reagent, 5X SSPE (900 mM NaCl; 50mM Sodium phosphate, pH7.7; and 5 mM EDTA), 0.1% SDS, and 200 µg/ml denatured salmon sperm DNA. After two hours, the filters were added to a fresh solution of the same composition with the addition of denatured radioactive hybridization probe. In this instance, the probes used were borage legumin cDNA (Fig. 16A), borage oleosin cDNA (Fig. 16B), and borage $\Delta 6$ -desaturase cDNA (pAN1, Example 2) (Fig. 16C). The borage legumin and oleosin cDNAs were isolated by EST cloning and identified by comparison to the GenBank database using the BLAST algorithm as described in Example 2. Loading variation was corrected by normalizing to

- 34 -

levels of borage EF1 α mRNA. EF1 α mRNA was identified by correlating to the corresponding cDNA obtained by the EST analysis described in Example 2. The filters were hybridized at 42°C for 12-20 hours, then washed as described above (except that the temperature was 65°C), air dried, and exposed to X-ray film.

As depicted in Figs. 15A and 15B, Δ 6-desaturase is expressed primarily in borage seed. Borage seeds reach maturation between 18-20 days post pollination (dpp). Δ 6-desaturase mRNA expression occurs throughout the time points collected (8-20 dpp), but appears maximal from 10-16 days post pollination. This expression profile is similar to that seen for borage oleosin and 12S seed storage protein mRNAs (Figs. 16A, 16B, and 16C).

-35-

EXAMPLE 5Isolation and Characterization of a Novel Oleosin cDNA

The oleosin cDNA (AtS21) was isolated by virtual subtraction screening of an *Arabidopsis* developing seed cDNA library using a random primed polymerase chain reaction (RP-PCR) cDNA probe derived from root tissue.

RNA PREPARATION

Arabidopsis thaliana Landsberg erecta plants were grown under continuous illumination in a vermiculite/soil mixture at ambient temperature (22°C). Siliques 2-5 days after flowering were dissected to separately collect developing seeds and silique coats. Inflorescences containing initial flower buds and fully opened flowers, leaves, and whole siliques one or three days after flowering were also collected. Roots were obtained from seedlings that had been grown in Gamborg B₅ liquid medium (GIBCO BRL) for two weeks. The seeds for root culture were previously sterilized with 50% bleach for five minutes and rinsed with water extensively. All tissues were frozen in liquid nitrogen and stored at -80°C until use. Total RNAs were isolated following a hot phenol/SDS extraction and LiCl precipitation protocol (Harris et al. 1978 *Biochem.* 17:3251-3256; Galau et al. 1981 *J. Biol. Chem.* 256:2551-2560). Poly A⁺ RNA was isolated using oligo dT column chromatography according to manufacturers' protocols (PHARMACIA or

- 36 -

STRATAGENE) or using oligotex-dT latex particles (QIAGEN).

Construction of tissue-specific cDNA libraries

Flower, one day silique, three day silique, leaf, root, and developing seed cDNA libraries were each constructed from 5 µg poly A+ RN using the ZAP cDNA synthesis kit (Stratagene). cDNAs were directionally cloned into the EcoRI and XhoI sites of pBluescript SK(-) in the λ-ZAPII vector (Short et al. 1988 *Nucleic Acids Res.* 16:7583-7600). Nonrecombinant phage plaques were identified by blue color development on NZY plates containing X-gal (5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside) and IPTG (isopropyl-1-thio-β-D-galactopyranoside). The nonrecombinant backgrounds for the flower, one day silique, three day silique, leaf, root, and developing seed cDNA libraries were 2.8%, 2%, 3.3%, 6.5%, 2.5%, and 1.9% respectively.

Random priming DNA labeling

The cDNA inserts of isolated clones (unhybridized cDNAs) were excised by EcoRI/XhoI double digestion and gel-purified for random priming labeling. Klenow reaction mixture contained 50 ng DNA templates, 10 mM Tris-HCl, pH 7.5, 5 mM MgCl₂, 7.5 mM DTT, 50 µM each of dCTP, dGTP, and dTTP, 10 µM hexamer random primers (Boehringer Mannheim), 50 µCi α-³²P-dATP, 3000 Ci/mmol, 10 mCi/ml (DuPont), and 5 units of DNA polymerase I Klenow fragment (New England

-37-

Biolabs). The reactions were carried out at 37°C for one hour. Aliquots of diluted reaction mixtures were used for TCA precipitation and alkaline denaturing gel analysis. Hybridization probes were labeled only with Klenow DNA polymerase and the unincorporated dNTPs were removed using Sephadex R G-50 spin columns (Boehringer Mannheim).

Random Primed PCR

Double-stranded cDNA was synthesized from poly A+ RNA isolated from *Arabidopsis* root tissue using the cDNA Synthesis System (GIBCO BRL) with oligo dT12-18 as primers. cDNAs longer than 300 bp were enriched by Sephacryl S-400 column chromatography (Stratagene). Fractionated cDNAs were used as templates for RP-PCR labeling. The reaction contained 10 mM Tris-HCl, pH 9.0, 50 mM KCl, 0.1% Triton X-100, 2 mM MgCl₂, 5 units Taq DNA polymerase (PROMEGA), 200 µM dCTP, cGTP, and dTTP, and different concentrations of hexamer random primers α-³²P dATP, 800 mCi/mMole, 10 mCi/ml (DuPont), and cold dATP in a final volume of 25 µl. After an initial 5 minutes at 95°C, different reactions were run through different programs to optimize RP-PCR cDNA conditions. Unless otherwise indicated, the following program was used for most RP-PCR cDNA probe labeling: 95°C/5 minutes, then 40 cycles of 95°C 30 seconds, 18°C/1 second, ramp to 30°C at a rate of 0.1°C/second. 72°C/1 minute. RP-PCR products were phenol/chloroform extracted and ethanol

- 38 -

precipitated or purified by passing through Sephadex G-50 spin columns (Boehringer Mannheim).

Clone blot virtual subtraction

Mass excision of λ -ZAP cDNA libraries was carried out by co-infecting XL1-Blue MRF' host cells with recombinant phage from the libraries and ExAssist helper phage (STRATAGENE). Excised phagemids were rescued by SOLR cells. Plasmid DNAs were prepared by boiling mini-prep method (Holmes et al. 1981 *Anal. Biochem.* 114:193-197) from randomly isolated clones. cDNA inserts were excised by EcoRI and XhoI double digestion, and resolved on 1% agarose gels. The DNAs were denatured in 0.5 N NaOH and 1.5 M NaCl for 45 minutes, neutralized in 0.5 M Tris-HCl, pH 8.0, and 1.5 M NaCl for 45 minutes, and then transferred by blotting to nylon membranes (Micron Separations, Inc.) in 10X SSC overnight. After one hour prehybridization at 65°C, root RP-cDNA probe was added to the same hybridization buffer containing 1% bovine albumin fraction V (Sigma), 1 mM EDTA, 0.5 M NaHPO₄, pH 7.2, 7% SDS. The hybridization continued for 24 hours at 65°C. The filters were washed in 0.5% bovine albumin, 1 mM EDTA, 40 mM NaHPO₄, pH 7.2, 5% SDS for ten minutes at room temperature, and 3 x 10 minutes in 1 mM EDTA, 40 mM NaHPO₄, pH 7.2, 1% SDS at 65°C. Autoradiographs were exposed to X-ray films (Kodak) for two to five days at -80°C.

Hybridization of resulting blots with root RP-PCR probes "virtually subtracted" seed cDNAs shared

- 39 -

with the root mRNA population. The remaining seed cDNAs representing putative seed-specific cDNAs, including those encoding oleosins, were sequenced by the cycle sequencing method, thereby identifying AtS21 as an oleosin cDNA clone.

Sequence analysis of AtS21

The oleosin cDNA is 834 bp long including an 18 bp long poly A tail (Fig. 4, SEQ ID NO:2). It has high homology to other oleosin genes from *Arabidopsis* as well as from other species. Recently, an identical oleosin gene has been reported (Zou, et al., 1996, *Plant Mol.Biol.* 31:429-433). The predicted protein is 191 amino acids long with a highly hydrophobic middle domain flanked by a hydrophilic domain on each side. The existence of two upstream in frame stop codons and the similarity to other oleosin genes indicate that this cDNA is full-length. Since there are two in frame stop codons just upstream of the first ATG, this cDNA is considered to be a full length cDNA (Figure 4, SEQ ID NO:2). The predicted protein has three distinctive domains based on the distribution of its amino acid residues. Both the N-terminal and C-terminal domains are rich in charged residues while the central domain is absolutely hydrophobic (Figure 5). As many as 20 leucine residues are located in the central domain and arranged as repeats with one leucine occurring every 7-10 residues. Other non-polar amino acid residues are also clustered in the central domain making this domain absolutely hydrophobic (Figure 6).

- 40 -

Extensive searches of different databases using both AtS21 cDNA and its predicted protein sequence identified oleosins from carrot, maize, cotton, rapeseed, *Arabidopsis*, and other plant species. The homology is mainly restricted to the central hydrophobic domain. Seven *Arabidopsis* oleosin sequences were found. AtS21 represents the same gene as Z54164 which has a few more bases in the 5' untranslated region. The seven *Arabidopsis* oleosin sequences available so far were aligned to each other (Figure 7). The result suggested that the seven sequences fall into three groups. The first group includes AtS21 (SEQ ID NO:5), X91918 (SEQ ID NO:6), and the partial sequence Z29859 (SEQ ID NO:7). Since X91918 (SEQ ID NO:6) has only its last residue different from AtS21 (SEQ ID NO:5), and since Z29859 (SEQ ID NO:7) has only three amino acid residues which are different from AtS21 (SEQ ID NO:5), all three sequences likely represent the same gene. The two sequences of the second group, X62352 (SEQ ID NO:8) and Ato13 (SEQ ID NO:9), are different in both sequence and length. Thus, there is no doubt that they represent two independent genes. Like the first group, the two sequences of the third group, X91956 (SEQ ID NO:10) and L40954 (SEQ ID NO:11), also have only three divergent residues which may be due to sequence errors. Thus, X91956 (SEQ ID NO:10) and L40954 (SEQ ID NO:11) likely represent the same gene. Unlike all the other oleosin sequences which were predicted from cDNA sequences, X62352 (SEQ ID NO:8)

was deduced from a genomic sequence (Van Rooijen et al. 1992 *Plant Mol. Biol.* 18:1177-1179). In conclusion, four different *Arabidopsis* oleosin genes have been identified so far, and they are conserved only in the middle of the hydrophobic domain.

Northern Analysis

In order to characterize the expression pattern of the native AtS21 gene, Northern analysis was performed as described in Example 4 except that the probe was the AtS21 cDNA (pAN1 insert) labeled with ^{32}P -dATP to a specific activity of 5×10^8 cpm/ug.

Results indicated that the AtS21 gene is strongly expressed in developing seeds and weakly expressed in silique coats (Figure 8A). A much larger transcript, which might represent unprocessed AtS21 pre-mRNA, was also detected in developing seed RNA. AtS21 was not detected in flower, leaf, root (Figure 8A), or one day silique RNAs. A different Northern analysis revealed that AtS21 is also strongly expressed in imbibed germinating seeds (Figs. 13A and 13B)

EXAMPLE 6**Characterization of Oleosin
Genomic Clones and Isolation of Oleosin Promoter**

Genomic clones were isolated by screening an *Arabidopsis* genomic DNA library using the full length cDNA (AtS21) as a probe. Two genomic clones were mapped by restriction enzyme digestion followed by Southern hybridization using the 5' half of the cDNA cleaved by SacI as a probe. A 2 kb SacI fragment was subcloned and sequenced (Fig. 9, SEQ ID NO:35). Two regions of the genomic clone are identical to the cDNA sequence. A 395 bp intron separates the two regions.

The copy number of AtS21 gene in the *Arabidopsis* genome was determined by genomic DNA Southern hybridization following digestion with the enzymes BamHI, EcoRI, HindIII, SacI and XbaI, using the full length cDNA as a probe (Figure 8B). A single band was detected in all the lanes except SacI digestion where two bands were detected. Since the cDNA probe has an internal SacI site, these results indicated that AtS21 is a single copy gene in the *Arabidopsis* genome. Since it has been known that *Arabidopsis* genome contains different isoforms of oleosin genes, this Southern analysis also demonstrates that the different oleosin isoforms of *Arabidopsis* are divergent at the DNA sequence level.

Two regions, separated by a 395 bp intron, of the genomic DNA fragment are identical to AtS21 cDNA sequence. Database searches using the 5' promoter sequence upstream of AtS21 cDNA sequence did

not identify any sequence with significant homology. Furthermore, the comparison of AtS21 promoter sequence with another *Arabidopsis* oleosin promoter isolated previously (Van Rooijen, et al., 1992) revealed little similarity. The AtS21 promoter sequence is rich in A/T bases, and contains as many as 44 direct repeats ranging from 10 bp to 14 bp with only one mismatch allowed. Two 14 bp direct repeats, and a putative ABA response element are underlined in Figure 9.

- 44 -

EXAMPLE 7

**Construction of AtS21
Promoter/GUS Gene Expression Cassette and Expression
Patterns in Transgenic Arabidopsis and Tobacco**

Construction of AtS21 promoter/GUS gene expression
cassette

The 1267 bp promoter fragment starting from the first G upstream of the ATG codon of the genomic DNA fragment was amplified using PCR and fused to the GUS reporter gene for analysis of its activity. The promoter fragment of the AtS21 genomic clone was amplified by PCR using the T7 primer GTAATACGACTCACTATAGGGC (SEQ ID NO:13) and the 21P primer GGGGATCCTATACTAAACTATAGAGTAAAGG (SEQ ID NO:14) complementary to the 5' untranslated region upstream of the first ATG codon (Figure 9). A BamHI cloning site was introduced by the 21P primer. The amplified fragment was cloned into the BamHI and SacI sites of pBluescript KS (Stratagene). Individual clones were sequenced to check possible PCR mutations as well as the orientation of their inserts. The correct clone was digested with BamHI and HindIII, and the excised promoter fragment (1.3 kb) was cloned into the corresponding sites of pBI101.1 (Jefferson, R.A. 1987a, *Plant Mol. Biol. Rep.* 5:387-405; Jefferson et al., 1987b, *EMBO J.* 6:3901-3907) upstream of the GUS gene. The resultant plasmid was designated pAN5 (Fig. 10). The AtS21 promoter/GUS construct (pAN5) was introduced into both tobacco (by the leaf disc method, Horsch et al., 1985; Bogue et al. 1990 *Mol. Gen. Gen.*

-45-

221:49-57) and *Arabidopsis* Colombia ecotype via vacuum infiltration as described by Bechtold, et al. (1993) *C.R. Acad. Sci. Paris*, 316:1194-1199. Seeds were sterilized and selected on media containing 50 μ g/ml kanamycin, 500 μ g/ml carbenicillin.

GUS activity assay: Expression patterns of the reporter GUS gene were revealed by histochemical staining (Jefferson, et al., 1987a, *Plant Mol. Biol. Rep.* 5:387-405). Different tissues were stained in substrate solution containing 2 mg/ml 5-bromo-4-chloro-3-indolyl- β -D-glucuronic acid (X-Gluc) (Research Organics, Inc.), 0.5 mM potassium ferrocyanide, and 0.5 mM potassium ferricyanide in 50 mM sodium phosphate buffer, pH 7.0 at 37°C overnight, and then dehydrated successively in 20%, 40% and 80% ethanol (Jefferson, et al., 1987). Photographs were taken using an Axiophot (Zeiss) compound microscope or Olympus SZH10 dissecting microscope. Slides were converted to digital images using a Spring/Scan 35LE slide scanner (Polaroid) and compiled using Adobe Photoshop™ 3.0.5 and Canvas™ 3.5.

GUS activities were quantitatively measured by fluorometry using 2 mM 4-MUG (4-methylumbelliferyl- β -D-glucuronide) as substrate (Jefferson, et al., 1987). Developing *Arabidopsis* seeds were staged according to their colors, and other plant tissues were collected and kept at -80°C until use. Plant tissues were ground in extraction buffer containing 50 mM sodium phosphate, pH 7.0, 10 mM EDTA, 10 mM β -mercaptoethanol, 0.1% Triton X-100, and 0.1% sodium

- 46 -

lauryl sarcosine. The tissue debris was removed by 5 minutes centrifugation in a microfuge. The supernatant was aliquoted and mixed with substrate and incubated at 37°C for 1 hour. Three replicas were assayed for each sample. The reactions were stopped by adding 4 volumes of 0.2 M sodium carbonate. Fluorescence was read using a TKO-100 DNA fluorometer (Hoefer Scientific Instruments). Protein concentrations of the extracts were determined by the Bradford method (Bio Rad).

Expression patterns of AtS21 promoter/GUS in transgenic *Arabidopsis* and tobacco

In *Arabidopsis*, GUS activity was detected in green seeds, and node regions where siliques, cauline leaves and branches join the inflorescence stem (Figures 11A and 11B). No GUS activity was detected in any leaf, root, flower, silique coat, or the internode regions of the inflorescence stem. Detailed studies of the GUS expression in developing seeds revealed that the AtS21 promoter was only active in green seeds in which the embryos had already developed beyond heart stage (Figures 11C and 11G). The youngest embryos showing GUS activity that could be detected by histochemical staining were at early torpedo stage. Interestingly, the staining was only restricted to the lower part of the embryo including hypocotyl and embryonic radical. No staining was detected in the young cotyledons (Figures 11D and 11E). Cotyledons began to be stained when the embryos

-47-

were at late torpedo or even early cotyledon stage (Figure 11F and 11H). Later, the entire embryos were stained, and the staining became more intense as the embryos matured (Figures 11I and 11J). It was also observed that GUS gene expression was restricted to the embryos. Seed coat and young endosperm were not stained (Figure 11C).

GUS activity was also detected in developing seedlings. Young seedlings of 3-5 days old were stained everywhere. Although some root hairs close to the hypocotyl were stained (Figure 11K), most of the newly formed structures such as root hairs, lateral root primordia and shoot apex were not stained (Figures 11L and 11N). Later, the staining was restricted to cotyledons and hypocotyls when lateral roots grew from the elongating embryonic root. The staining on embryonic roots disappeared. No staining was observed on newly formed lateral roots, true leaves nor trichomes on true leaves (Figures 11M and 11N).

AtS21 promoter/GUS expression patterns in tobacco are basically the same as in *Arabidopsis*. GUS activity was only detected in late stage seeds and different node regions of mature plants. In germinating seeds, strong staining was detected throughout the entire embryos as soon as one hour after they were dissected from imbibed seeds. Mature endosperm, which *Arabidopsis* seeds do not have, but not seed coat was also stained (Figure 12A). The root tips of some young seedlings of one transgenic line

- 48 -

were not stained (Figure 12B). Otherwise, GUS expression patterns in developing tobacco seedlings were the same as in *Arabidopsis* seedlings (Figures 12B, 12C, and 12D). Newly formed structures such as lateral roots and true leaves were not stained.

AtS21 mRNA levels in developing seedlings

Since the observed strong activities of AtS21 promoter/GUS in both *Arabidopsis* and tobacco seedlings are not consistent with the seed-specific expression of oleosin genes, Northern analysis was carried out to determine if AtS21 mRNA was present in developing seedlings where the GUS activity was so strong. RNAs prepared from seedlings at different stages from 24 hours to 12 days were analyzed by Northern hybridization using AtS21 cDNA as the probe. Surprisingly, AtS21 mRNA was detected at a high level comparable to that in developing seeds in 24-48 hour imbibed seeds. The mRNA level dropped dramatically when young seedlings first emerged at 74 hours (Figures 13A and 13B). In 96 hour and older seedlings, no signal was detected even with a longer exposure (Figure 13B). The loadings of RNA samples were checked by hybridizing the same blot with a tubulin gene probe (Figure 13C) which was isolated and identified by EST analysis as described in Example 2. Since AtS21 mRNA was so abundant in seeds, residual AtS21 probes remained on the blot even after extensive stripping. These results indicated that AtS21 mRNA detected in imbibed seeds and very young seedlings are

- 49 -

the carry-over of AtS21 mRNA from dry seeds. It has recently been reported that an oleosin Atol2 mRNA (identical to AtS21) is most abundant in dry seeds (Kirik, et al., 1996 *Plant Mol. Biol.* 31(2):413-417.) Similarly, the strong GUS activities in seedlings were most likely due to the carry-over of both β -glucuronidase protein and the *de novo* synthesis of β -glucuronidase from its mRNA carried over from the dry seed stage.

EXAMPLE 8Activity comparison between the
AtS21 promoter and the 35S promoter

The GUS activities in transgenic *Arabidopsis* developing seeds expressed by the AtS21 promoter were compared with those expressed by the 35S promoter in the construct pBI221 (Jefferson et al. *EMBO J.* 6:3901-3907). The seeds were staged according to their colors (Table 2). The earliest stage was from globular to late heart stage when the seeds were still white but large enough to be dissected from the siliques. AtS21 promoter activity was detected at a level about three times lower than that of the 35S promoter at this stage. 35S promoter activity remained at the same low level throughout the entire embryo development. In contrast, AtS21 promoter activity increased quickly as the embryos passed torpedo stage and reached the highest level of 25.25 pmole 4-MU/min. μ g protein at mature stage (Figure 5-8). The peak activity of the AtS21 promoter is as much as 210 times higher than its lowest activity at globular to heart stage, and is close to 100 times higher than the 35S promoter activity at the same stage (Table 2). The activity levels of the AtS21 promoter are similar to those of another *Arabidopsis* oleosin promoter expressed in *Brassica napus* (Plant et al. 1994, *Plant mol. Biol.* 25:193-205. AtS21 promoter activity was also detected at background level in leaf. The high standard deviation, higher than the average itself, indicated that the GUS activity was

- 51 -

only detected in the leaves of some lines (Table 2). On the other hand, 35S promoter activity in leaf was more than 20 times higher than that in seed. The side by side comparisons of activities between AtS21 promoter and 35S promoter is shown in Figure 14.

Although the AtS21 promoter activity was about 3 times lower in dry seed of tobacco than in *Arabidopsis* dry seed, the absolute GUS activity was still higher than that expressed by the 35S promoter in *Arabidopsis* leaf (Table 2). No detectable AtS21 promoter activity was observed in tobacco leaf (Figure 14).

Comparison of the AtS21 promoter versus the 35S promoter revealed that the latter is not a good promoter to express genes at high levels in developing seeds. Because of its consistent low activities throughout the entire embryo development period, 35S promoter is useful for consistent low level expression of target genes. On the other hand, the AtS21 promoter is a very strong promoter that can be used to express genes starting from heart stage embryos and accumulating until the dry seed stage. The 35S promoter, although not efficient, is better than the AtS21 promoter in expressing genes in embryos prior to heart stage.

35 25 20 15 10 5 1

TABLE 2
GUS ACTIVITIES OF AtS21 and 35S PROMOTER/GUS CONSTRUCTS

COLOR STAGE	WHITE G-H	WHITE/YELLOW H-T	YELLOW T-C	LIGHT GREEN EARLY C	DARK GREEN LATE C	GREEN/YELLOW/BROWN MATURE	BROWN DRY SEED	LEAF
AtS21	0.12±0.17	1.35±1.57	6.77±1.25	18.99±3.75	21.85±4.45	25.25±4.64	24.38±10.85	0.08±0.
35S	0.30±0.06	0.25±0.08	0.29±0.04	0.28±0.03	0.33±0.06	0.26±0.04	0.31±0.02	6.56±0.
AtS21	(In tobacco)						8.81±0.21	0.01±0.

Abbreviations: G, globular stage; H, heart stage; T, torpedo stage; C, cotyledon stage. The GUS activities are in pmole 4-MU/µg protein.min. For AtS21 promoter the numbers are the average of five independent lines with standard deviations. Three repeats were assayed for each line. For 35S promoter the numbers are the average of three repeats of the same line with standard deviations.

EXAMPLE 9

Expression of the Borage Δ^6 -Desaturase Gene Under the Control of the AtS21 Promoter and Comparison to Expression Under the Control of the CaMV 35S Promoter

In order to create an expression construct with the AtS21 promoter driving expression of the borage Δ^6 -desaturase gene, the GUS coding fragment from pAN5 was removed by digestion with SmaI and EcoICR I. The cDNA insert of pAN1 (Example 2) was then excised by first digesting with XhoI (and filling in the residual overhang as above), and then digesting with SmaI. The resulting fragment was used to replace the excised portion of pAN5, yielding pAN3.

After transformation of tobacco and *Arabidopsis* following the methods of Example 7, levels of Δ^6 -desaturase activity were monitored by assaying the corresponding fatty acid methyl esters of its reaction products, γ -linolenic acid (GLA) and octadecatetraenoic acid (OTA) using the methods referred to in Example 3. The GLA and OTA levels (Table 3) of the transgenic seeds ranged up to 6.7% of C18 fatty acids (Mean = 3.1%) and 2.8% (Mean = 1.1%), respectively. No GLA or OTA was detected in the leaves of these plants. In comparison, CaMV 35 S promoter/ Δ^6 -desaturase transgenic plants produced GLA levels in seeds ranging up to 3.1% of C18 fatty acids (Mean = 1.3%) and no measurable OTA in seeds.

TABLE 3
EXPRESSION OF THE BORAGE Δ⁶-DESATURASE IN TRANSGENIC PLANTS

PROMOTER	PLANT	SEED			LEAF		
		GLA*	RANGE	OTA*	RANGE	OTA	RANGE
Cauliflower mosaic virus 35S	tobacco	1.3	0.7-3.1	n.d	20	19-22	9.7 8-11
Arabidopsis oleosin	Arabidopsis	3.1	0-6.7	1.1	0-2.8	n.d.	n.d

*mean value expressed as the percent of the C₁₈ fatty acids
n.d. not detected

EXAMPLE 10**Transformation of Oilseed Rape With an Expression Cassette Which Comprises the Oleosin 5' Regulatory Region Linked to the Borage Delta 6-Desaturase Gene**

Oilseed rape, Cv. Westar, was transformed with the strain of *Agrobacterium tumefaciens* EHA105 containing the plasmid pAN3 (i.e. the borage $\Delta 6$ -desaturase gene under the control of the *Arabidopsis* oleosin promoter-Example 9).

Terminal internodes of Westar were co-cultivated for 2-3 days with induced *Agrobacterium tumefaciens* strain EHA105 (Alt-Moerbe et al. 1988 *Mol. Gen. Genet.* 213:1-8; James et al. 1993 *Plant Cell Reports* 12:559-563), then transferred onto regeneration medium (Boulter et al. 1990 *Plant Science* 70:91-99; Fry et al. 1987 *Plant Cell Reports* 6:321-325). The regenerated shoots were transferred to growth medium (Pelletier et al. 1983 *Mol. Gen. Menet.* 191:244-250), and a polymerase chain reaction (PCR) test was performed on leaf fragments to assess the presence of the gene.

DNA was isolated from the leaves according to the protocol of KM Haymes et al. (1996) *Plant Molecular Biology Reporter* 14(3):280-284, and resuspended in 100 μ l of water, without RNase treatment. 5 μ l of extract were used for the PCR reaction, in a final volume of 50 μ l. The reaction was performed in a Perkin-Elmer 9600 thermocycler, with the following cycles:

-56-

1 cycle: 95°C, 5 minutes
30 cycles: 95°C, 45 sec; 52°C, 45 sec
72°C, 1 minute
1 cycle: 72°C, 5 minutes

and the following primers (derived from near the metal box regions, as indicated in Fig. 1, SEQ. NO.:1):

5' TGG AAA TGG AAC CAT AA 3'

5' GGA AAC AAA TGA TGC TC 3'

Amplification of the DNA revealed the expected 549 base pair PCR fragment (Figure 17).

The positive shoots were transferred to elongation medium, then to rooting medium (DeBlock et al 1989 *Plant Physiol.* 91:694-701). Shoots with a well-developed root system were transferred to the greenhouse. When plants were well developed, leaves were collected for Southern analysis in order to assess gene copy number.

Genomic DNA was extracted according to the procedure of Bouchez et al. (1996) *Plant Molecular Biology Reporter* 14:115-123, digested with the restriction enzymes *Bgl* I and/or *Cla* I, electrophoretically separated on agarose gel (Maniatis et al. 1982, in *Molecular Cloning; a Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor/NY), and prepared for transfer to nylon membranes (Nytran membrane, Schleicher & Schuell) according to the instructions of the manufacturer. DNA was then transferred to membranes overnight by capillary action using 20XSSC (Maniatis et al. 1982).

- 57 -

Following transfer, the membranes were crosslinked by UV (Stratagene) for 30 seconds and pre-hybridized for 1 hour at 65°C in 15 ml of a solution containing 6XSSC, 0.5%SDS and 2.25% w/w dehydrated skim milk in glass vials in hybridization oven (Appligene). The membranes were hybridized overnight in the same solution containing a denatured hybridization probe radiolabelled with ^{32}P to a specific activity of 10^8 cpm/ μg by the random primer method (with the Ready-To-Go kit obtained from Pharmacia). The probe represents a PCR fragment of the borage delta 6-desaturase gene (obtained in the conditions and with the primers detailed above). After hybridization, the filters were washed at 65°C in 2XSSC, 0.1% SDS for 15 minutes, and 0.2XSSC, 0.1%SDS for 15 minutes. The membranes were then wrapped in Saran-Wrap and exposed to Kodak XAR film using an intensifying screen at -70°C in a light-proof cassette. Exposure time was generally 3 days.

The results obtained confirm the presence of the gene. According to the gene construct, the number of bands in each lane of DNA digested by *Bgl* I or *Cla* I represents the number of delta 6-desaturase genes present in the genomic DNA of the plant. The digestion with *Bgl* I and *Cla* I together generates a fragment of 3435 bp.

The term "comprises" or "comprising" is defined as specifying the presence of the stated features, integers, steps, or components as referred to in the claims, but does not preclude the presence or addition of one or more other features, integers, steps, components, or groups thereof.

WO 98/45461

PCT/US98/07179

-58-

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Rhone Poulenc Agro
Thomas, Terry L.
Li, Zhongsen

(ii) TITLE OF INVENTION: AN OLEOSIN 5' REGULATORY REGION FOR THE
MODIFICATION OF PLANT SEED LIPID COMPOSITION

(iii) NUMBER OF SEQUENCES: 35

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Scully, Scott, Murphy & Presser
(B) STREET: 400 Garden City Plaza
(C) CITY: Garden City
(D) STATE: New York
(E) COUNTRY: USA
(F) ZIP: 11530

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: 08/831,575
(B) FILING DATE: 9 April 1997
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: DiGiglio, Frank S.
(B) REGISTRATION NUMBER: 31,346
(C) REFERENCE/DOCKET NUMBER: 10203

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (516) 742-4343
(B) TELEFAX: (516) 742-4366

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1684 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 43..1387

WO 98/45461

PCT/US98/07179

-59-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATATCTGCCT ACCCTCCCAA AGAGAGTAGT CATTTTTCAT CA	ATG GCT GCT CAA	54
	Met Ala Ala Gln	
	1	
ATC AAG AAA TAC ATT ACC TCA GAT GAA CTC AAG AAC CAC GAT AAA CCC		102
Ile Lys Lys Tyr Ile Thr Ser Asp Glu Leu Lys Asn His Asp Lys Pro		
5 10 15 20		
GGA GAT CTA TGG ATC TCG ATT CAA GGG AAA GCC TAT GAT GTT TCG GAT		150
Gly Asp Leu Trp Ile Ser Ile Gln Gly Lys Ala Tyr Asp Val Ser Asp		
25 30 35		
TGG GTG AAA GAC CAT CCA GGT GGC AGC TTT CCC TTG AAG AGT CTT GCT		198
Trp Val Lys Asp His Pro Gly Gly Ser Phe Pro Leu Lys Ser Leu Ala		
40 45 50		
GGT CAA GAG GTA ACT GAT GCA TTT GTT GCA TTC CAT CCT GCC TCT ACA		246
Gly Gln Glu Val Thr Asp Ala Phe Val Ala Phe His Pro Ala Ser Thr		
55 60 65		
TGG AAG AAT CTT GAT AAG TTT TTC ACT GGG TAT TAT CTT AAA GAT TAC		294
Trp Lys Asn Leu Asp Lys Phe Phe Thr Gly Tyr Tyr Leu Lys Asp Tyr		
70 75 80		
TCT GTT TCT GAG GTT TCT AAA GAT TAT AGG AAG CTT GTG TTT GAG TTT		342
Ser Val Ser Glu Val Ser Lys Asp Tyr Arg Lys Leu Val Phe Glu Phe		
85 90 95 100		
TCT AAA ATG GGT TTG TAT GAC AAA AAA GGT CAT ATT ATG TTT GCA ACT		390
Ser Lys Met Gly Leu Tyr Asp Lys Lys Gly His Ile Met Phe Ala Thr		
105 110 115		
TTG TGC TTT ATA GCA ATG CTG TTT GCT ATG AGT GTT TAT GGG GTT TTG		438
Leu Cys Phe Ile Ala Met Leu Phe Ala Met Ser Val Tyr Gly Val Leu		
120 125 130		
TTT TGT GAG GGT GTT TTG GTA CAT TTG TTT TCT GGG TGT TTG ATG GGG		486
Phe Cys Glu Gly Val Leu Val His Leu Phe Ser Gly Cys Leu Met Gly		
135 140 145		
TTT CTT TGG ATT CAG AGT GGT TGG ATT GGA CAT GAT GCT GGG CAT TAT		534
Phe Leu Trp Ile Gln Ser Gly Trp Ile Gly His Asp Ala Gly His Tyr		
150 155 160		
ATG GTA GTG TCT GAT TCA AGG CTT AAT AAG TTT ATG GGT ATT TTT GCT		582
Met Val Val Ser Asp Ser Arg Leu Asn Lys Phe Met Gly Ile Phe Ala		
165 170 175 180		
GCA AAT TGT CTT TCA GGA ATA AGT ATT GGT TGG TGG AAA TGG AAC CAT		630
Ala Asn Cys Leu Ser Gly Ile Ser Ile Gly Trp Trp Lys Trp Asn His		
185 190 195		

WO 98/45461

PCT/US98/07179

-60-

AAT Asn	GCA Ala	CAT His	CAC His 200	ATT Ile	GCC Ala	TGT Cys	AAT Asn	AGC Ser 205	CTT Leu	GAA Glu	TAT Tyr	GAC Asp	CCT Pro 210	GAT Asp	TTA Leu	678
CAA Gln	TAT Tyr	ATA Ile 215	CCA Pro	TTC Phe	CTT Leu	GTT Val	GTG Val 220	TCT Ser	TCC Ser	AAG Lys	TTT Phe	TTT Phe 225	GGT Gly	TCA Ser	CTC Leu	726
ACC Thr	TCT Ser 230	CAT His	TTC Phe	TAT Tyr	GAG Glu	AAA Lys 235	AGG Arg	TTG Leu	ACT Thr	TTT Phe	GAC Asp 240	TCT Ser	TTA Leu	TCA Ser	AGA Arg	774
TTC Phe 245	TTT Phe	GTA Val	AGT Ser	TAT Tyr	CAA Gln 250	CAT His	TGG Trp	ACA Thr	TTT Phe	TAC Tyr 255	CCT Pro	ATT Ile	ATG Met	TGT Cys	GCT Ala 260	822
GCT Ala	AGG Arg	CTC Leu	AAT Asn	ATG Met 265	TAT Tyr	GTA Val	CAA Gln	TCT Ser	CTC Leu 270	ATA Ile	ATG Met	TTG Leu	TTG Leu	ACC Thr 275	AAG Lys	870
AGA Arg	AAT Asn	GTG Val	TCC Ser 280	TAT Tyr	CGA Arg	GCT Ala	CAG Gln	GAA Glu 285	CTC Leu	TTG Leu	GGA Gly	TGC Cys	CTA Leu 290	GTG Val	TTC Phe	918
TCG Ser	ATT Ile	TGG Trp 295	TAC Tyr	CCG Pro	TTG Leu	CTT Leu	GTT Val 300	TCT Ser	TGT Cys	TTG Leu	CCT Pro	AAT Asn 305	TGG Trp	GGT Gly	GAA Glu	966
AGA Arg	ATT Ile 310	ATG Met	TTT Phe	GTT Val	ATT Ile	GCA Ala 315	AGT Ser	TTA Leu	TCA Ser	GTG Val	ACT Thr 320	GGA Gly	ATG Met	CAA Gln	CAA Gln	1014
GTT Val 325	CAG Gln	TTC Phe	TCC Ser	TTG Leu	AAC Asn 330	CAC His	TTC Phe	TCT Ser	TCA Ser	AGT Ser 335	GTT Val	TAT Tyr	GTT Val	GGA Gly	AAG Lys 340	1062
CCT Pro	AAA Lys	GGG Gly	AAT Asn	AAT Asn 345	TGG Trp	TTT Phe	GAG Glu	AAA Lys	CAA Gln 350	ACG Thr	GAT Asp	GGG Gly	ACA Thr	CTT Leu 355	GAC Asp	1110
ATT Ile	TCT Ser	TGT Cys	CCT Pro 360	CCT Pro	TGG Trp	ATG Met	GAT Asp	TGG Trp 365	TTT Phe	CAT His	GGT Gly	GGA Gly	TTG Leu 370	CAA Gln	TTC Phe	1158
CAA Gln	ATT Ile	GAG Glu 375	CAT His	CAT His	TTG Leu	TTT Phe	CCC Pro 380	AAG Lys	ATG Met	CCT Pro	AGA Arg	TGC Cys 385	AAC Asn	CTT Leu	AGG Arg	1206
AAA Lys	ATC Ile 390	TCG Ser	CCC Pro	TAC Tyr	GTG Val	ATC Ile 395	GAG Glu	TTA Leu	TGC Cys	AAG Lys	AAA Lys 400	CAT His	AAT Asn	TTG Leu	CCT Pro	1254
TAC Tyr 405	AAT Asn	TAT Tyr	GCA Ala	TCT Ser	TTC Phe 410	TCC Ser	AAG Lys	GCC Ala	AAT Asn	GAA Glu 415	ATG Met	ACA Thr	CTC Leu	AGA Arg	ACA Thr 420	1302

-61-

TTG	AGG	AAC	ACA	GCA	TTG	CAG	GCT	AGG	GAT	ATA	ACC	AAG	CCG	CTC	CCG		
Leu	Arg	Asn	Thr	Ala	Leu	Gln	Ala	Arg	Asp	Ile	Thr	Lys	Pro	Leu	Pro		1350
				425					430					435			

AAG AAT TTG GTA TGG GAA GCT CTT CAC ACT CAT GGT T AAAATTACCC
Lys Asn Leu Val Trp Glu Ala Leu His Thr His Gly
440 445

TTAGTTCATG TAATAATTTG AGATTATGTA TCTCCTATGT TTGTGTCTTG TCTTGGTTCT 1457

ACTTGTTGGA GTCATTGCAA CTTGTCTTTT ATGGTTTATT AGATGTTTTT TAATATATTT 1517

TAGAGGTTTT GCTTTCATCT CCATTATTGA TGAATAAGGA GTTGCATATT GTCAATTGTT 1577

GTGCTCAATA TCTGATATTT TGAATGTAC TTTGTACCAC GTGGTTTTCA GTTGAAGCTC 1637

ATGTGTACTT CTATAGACTT TGTTTAAATG GTTATGTCAT GTTATTT 1684

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 135 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Ala Ala Gln Ile Lys Lys Tyr Ile Thr Ser Asp Glu Leu Lys Asn
1 5 10 15

His Asp Lys Pro Gly Asp Leu Trp Ile Ser Ile Gln Gly Lys Ala Tyr
20 25 30

Asp Val Ser Asp Trp Val Lys Asp His Pro Gly Gly Ser Phe Pro Leu
35 40
Lys Ser Leu Ala Gly Gln Glu Val Thr Asp Ala Phe Val Ala Phe His
50 55 60

Pro Ala Ser Thr Trp Lys Asn Leu Asp Lys Phe Phe Thr Gly Tyr Tyr
65 70 75 80

Leu Lys Asp Tyr Ser Val Ser Glu Val Ser Lys Asp Tyr Arg Lys Leu
85 90 95

Val Phe Glu Phe Ser Lys Met Gly Leu Tyr Asp Lys Lys Gly His Ile
100 105 110

Met Phe Ala Thr Leu Cys Phe Ile Ala Met Leu Phe Ala Met Ser Val
115 120 125

Tyr Gly Val Leu Phe Cys Glu
130 135

WO 98/45461

PCT/US98/07179

-62-

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 834 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
 (B) LOCATION: 31..603

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TTAGCCTTTA CTCTATAGTT TTAGATAGAC ATG GCG AAT GTG GAT CGT GAT CGG	54
Met Ala Asn Val Asp Arg Asp Arg	
140	
CGT GTG CAT GTA GAC CGT ACT GAC AAA CGT GTT CAT CAG CCA AAC TAC	102
Arg Val His Val Asp Arg Thr Asp Lys Arg Val His Gln Pro Asn Tyr	
145 150 155	
GAA GAT GAT GTC GGT TTT GGT GGC TAT GGC GGT TAT GGT GCT GGT TCT	150
Glu Asp Asp Val Gly Phe Gly Gly Tyr Gly Gly Tyr Gly Ala Gly Ser	
160 165 170 175	
GAT TAT AAG AGT CGC GGC CCC TCC ACT AAC CAA ATC TTG GCA CTT ATA	198
Asp Tyr Lys Ser Arg Gly Pro Ser Thr Asn Gln Ile Leu Ala Leu Ile	
180 185 190	
GCA GGA GTT CCC ATT GGT GGC ACA CTG CTA ACC CTA GCT GGA CTC ACT	246
Ala Gly Val Pro Ile Gly Gly Thr Leu Leu Thr Leu Ala Gly Leu Thr	
195 200 205	
CTA GCC GGT TCG GTG ATC GGC TTG CTA GTC TCC ATA CCC CTC TTC CTC	294
Leu Ala Gly Ser Val Ile Gly Leu Leu Val Ser Ile Pro Leu Phe Leu	
210 215 220	
CTC TTC AGT CCG GTG ATA GTC CCG GCG GCT CTC ACT ATT GGG CTT GCT	342
Leu Phe Ser Pro Val Ile Val Pro Ala Ala Leu Thr Ile Gly Leu Ala	
225 230 235	
GTG ACG GGA ATC TTG GCT TCT GGT TTG TTT GGG TTG ACG GGT CTG AGC	390
Val Thr Gly Ile Leu Ala Ser Gly Leu Phe Gly Leu Thr Gly Leu Ser	
240 245 250 255	
TCG GTC TCG TGG GTC CTC AAC TAC CTC CGT GGG ACG AGT GAT ACA GTG	438
Ser Val Ser Trp Val Leu Asn Tyr Leu Arg Gly Thr Ser Asp Thr Val	
260 265 270	

WO 98/45461

PCT/US98/07179

-63-

CCA GAG CAA TTG GAC TAC GCT AAA CGG CGT ATG GCT GAT GCG GTA GGC Pro Glu Gln Leu Asp Tyr Ala Lys Arg Arg Met Ala Asp Ala Val Gly 275 280 285	486
TAT GCT GGT ATG AAG GGA AAA GAG ATG GGT CAG TAT GTG CAA GAT AAG Tyr Ala Gly Met Lys Gly Lys Glu Met Gly Gln Tyr Val Gln Asp Lys 290 295 300	534
GCT CAT GAG GCT CGT GAG ACT GAG TTC ATG ACT GAG ACC CAT GAG CCG Ala His Glu Ala Arg Glu Thr Glu Phe Met Thr Glu Thr His Glu Pro 305 310 315	582
GGT AAG GCC AGG AGA GGC TCA TAAGCTAATA TAAATTGCGG GAGTCAGTTG Gly Lys Ala Arg Arg Gly Ser 320 325	633
GAAACGCGAT AAATGTAGTT TTA CTTTCTCTT TTAAGAATAT CTTTGTCTAT ATATGTGTTC GTTCGTTTTG TCTTGTCCAA ATAAAAATCC TTGTTAGTGA AATAAGAAAT GAAATAAATA TGTTTTCTTT TTTGAGATAA CCAGAAATCT CATACTATTT TCTAAAAAAA AAAAAAAAAA A	693 753 813 834

(2) INFORMATION FOR 'SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 191 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Ala Asn Val Asp Arg Asp Arg Arg Val His Val Asp Arg Thr Asp 1 5 10 15
Lys Arg Val His Gln Pro Asn Tyr Glu Asp Asp Val Gly Phe Gly Gly 20 25 30
Tyr Gly Gly Tyr Gly Ala Gly Ser Asp Tyr Lys Ser Arg Gly Pro Ser 35 40 45
Thr Asn Gln Ile Leu Ala Leu Ile Ala Gly Val Pro Ile Gly Gly Thr 50 55 60
Leu Leu Thr Leu Ala Gly Leu Thr Leu Ala Gly Ser Val Ile Gly Leu 65 70 75 80
Leu Val Ser Ile Pro Leu Phe Leu Leu Phe Ser Pro Val Ile Val Pro 85 90 95

WO 98/45461

PCT/US98/07179

-64-

Ala	Ala	Leu	Thr	Ile	Gly	Leu	Ala	Val	Thr	Gly	Ile	Leu	Ala	Ser	Gly
			100					105					110		
Leu	Phe	Gly	Leu	Thr	Gly	Leu	Ser	Ser	Val	Ser	Trp	Val	Leu	Asn	Tyr
		115					120					125			
Leu	Arg	Gly	Thr	Ser	Asp	Thr	Val	Pro	Glu	Gln	Leu	Asp	Tyr	Ala	Lys
		130				135					140				
Arg	Arg	Met	Ala	Asp	Ala	Val	Gly	Tyr	Ala	Gly	Met	Lys	Gly	Lys	Glu
145					150					155					160
Met	Gly	Gln	Tyr	Val	Gln	Asp	Lys	Ala	His	Glu	Ala	Arg	Glu	Thr	Glu
				165					170					175	
Phe	Met	Thr	Glu	Thr	His	Glu	Pro	Gly	Lys	Ala	Arg	Arg	Gly	Ser	
			180					185					190		

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 191 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met	Ala	Asn	Val	Asp	Arg	Asp	Arg	Arg	Val	His	Val	Asp	Arg	Thr	Asp
1				5					10					15	
Lys	Arg	Val	His	Gln	Pro	Asn	Tyr	Glu	Asp	Asp	Val	Gly	Phe	Gly	Gly
			20					25					30		
Thr	Gly	Gly	Thr	Gly	Ala	Gly	Ser	Asp	Tyr	Lys	Ser	Arg	Gly	Pro	Ser
		35				40						45			
Thr	Asn	Gln	Ile	Leu	Ala	Leu	Ile	Ala	Gly	Val	Pro	Ile	Gly	Gly	Thr
	50					55				60					
Leu	Ile	Thr	Leu	Ala	Gly	Leu	Thr	Leu	Ala	Gly	Ser	Val	Ile	Gly	Leu
65					70				75					80	
Leu	Val	Ser	Ile	Pro	Leu	Phe	Leu	Ile	Phe	Ser	Pro	Val	Ile	Val	Pro
			85					90					95		
Ala	Ala	Leu	Thr	Ile	Gly	Leu	Ala	Val	Thr	Gly	Ile	Leu	Ala	Ser	Gly
			100					105					110		
Leu	Phe	Gly	Leu	Thr	Gly	Leu	Ser	Ser	Val	Ser	Trp	Val	Leu	Asn	Tyr
		115					120					125			

WO 98/45461

PCT/US98/07179

-65-

Leu Arg Gly Thr Ser Asp Thr Val Pro Glu Gln Leu Asp Tyr Ala Lys
 130 135 140
 Arg Arg Met Ala Asp Ala Val Gly Tyr Ala Gly Met Lys Gly Lys Glu
 145 150 155 160
 Met Gly Gln Tyr Val Gln Asp Lys Ala His Glu Ala Arg Glu Thr Glu
 165 170 175
 Phe Met Thr Glu Thr His Glu Pro Gly Lys Ala Arg Arg Gly Ser
 180 185 190

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 191 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Ala Asn Val Asp Arg Asp Arg Arg Val His Val Asp Arg Thr Asp
 1 5 10 15
 Lys Arg Val His Gln Pro Asn Tyr Glu Asp Asp Val Gly Phe Gly Gly
 20 25 30
 Thr Gly Gly Thr Gly Ala Gly Ser Asp Tyr Lys Ser Arg Gly Pro Ser
 35 40 45
 Thr Asn Gln Ile Leu Ala Leu Ile Ala Gly Val Pro Ile Gly Gly Thr
 50 55 60
 Leu Ile Thr Leu Ala Gly Leu Thr Leu Ala Gly Ser Val Ile Gly Leu
 65 70 75 80
 Leu Val Ser Ile Pro Leu Phe Leu Ile Phe Ser Pro Val Ile Val Pro
 85 90 95
 Ala Ala Leu Thr Ile Gly Leu Ala Val Thr Gly Ile Leu Ala Ser Gly
 100 105 110
 Leu Phe Gly Leu Thr Gly Leu Ser Ser Val Ser Trp Val Leu Asn Tyr
 115 120 125
 Leu Arg Gly Thr Ser Asp Thr Val Pro Glu Gln Leu Asp Tyr Ala Lys
 130 135 140
 Arg Arg Met Ala Asp Ala Val Gly Tyr Ala Gly Met Lys Gly Lys Glu
 145 150 155 160

WO 98/45461

PCT/US98/07179

-66-

Met	Gly	Gln	Tyr	Val	Gln	Asp	Lys	Ala	His	Glu	Ala	Arg	Glu	Thr	Glu
				165					170					175	
Phe	Met	Thr	Glu	Thr	His	Glu	Pro	Gly	Lys	Ala	Arg	Arg	Gly	Pro	
			180					185					190		

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 78 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Phe	Gly	Leu	Thr	Gly	Leu	Ser	Ser	Val	Ser	Trp	Val	Leu	Gln	Leu	Pro
1				5					10					15	
Pro	Trp	Ala	Ser	Asp	Thr	Val	Pro	Glu	Gln	Val	Asp	Tyr	Ala	Lys	Arg
			20					25					30		
Arg	Met	Ala	Asp	Ala	Val	Gly	Tyr	Ala	Gly	Met	Lys	Gly	Lys	Glu	Met
		35				40						45			
Gly	Gln	Tyr	Val	Gln	Asp	Lys	Ala	His	Glu	Ala	Arg	Glu	Thr	Glu	Phe
	50				55						60				
Met	Thr	Glu	Thr	His	Glu	Pro	Gly	Lys	Ala	Arg	Arg	Gly	Ser		
65					70					75					

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 173 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met	Ala	Asp	Thr	Ala	Arg	Gly	Thr	His	His	Asp	Ile	Ile	Gly	Arg	Asp
1				5					10					15	
Gln	Tyr	Pro	Met	Met	Gly	Arg	Asp	Arg	Asp	Gln	Tyr	Gln	Met	Ser	Gly
			20					25					30		
Arg	Gly	Ser	Asp	Tyr	Ser	Lys	Ser	Arg	Gln	Ile	Ala	Lys	Ala	Ala	Thr
		35				40						45			

WO 98/45461

PCT/US98/07179

-67-

Ala Val Thr Ala Gly Gly Ser Leu Leu Val Leu Ser Ser Leu Thr Leu
50 55 60
Val Gly Thr Val Leu Ala Leu Thr Val Ala Thr Pro Leu Leu Val Leu
65 70 75 80
Phe Ser Pro Ile Leu Val Pro Ala Leu Ile Thr Val Ala Leu Leu Ile
85 90 95
Thr Gly Phe Leu Ser Ser Gly Gly Phe Gly Ile Ala Ala Ile Thr Val
100 105 110
Phe Ser Trp Ile Tyr Lys Tyr Ala Thr Gly Glu His Pro Gln Gly Ser
115 120 125
Asp Lys Leu Asp Ser Ala Arg Met Lys Leu Gly Ser Lys Ala Gln Asp
130 135 140
Leu Lys Asp Arg Ala Gln Tyr Tyr Gly Gln Gln His Thr Gly Gly Glu
145 150 155 160
His Asp Arg Asp Arg Thr Arg Gly Gly Gln His Thr Thr
165 170

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 141 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met Ala Asp Gln Thr Arg Thr His His Glu Met Ile Ser Arg Asp Ser
1 5 10 15
Thr Gln Glu Ala His Pro Lys Ala Arg Gln Trp Val Lys Ala Ala Thr
20 25 30
Ala Val Thr Ala Gly Gly Ser Leu Leu Val Leu Ser Gln Leu Thr Leu
35 40 45
Ala Gly Thr Val Ile Ala Leu Thr Val Ala Thr Pro Leu Leu Val Ile
50 55 60
Phe Ser Pro Val Leu Val Pro Ala Val Val Thr Val Ala Leu Ile Ile
65 70 75 80
Thr Gly Phe Leu Ala Ser Gly Gly Phe Gly Ile Ala Ala Ile Thr Ala
85 90 95

WO 98/45461

PCT/US98/07179

-68-

Phe	Ser	Trp	Leu	Tyr	Arg	His	Trp	Thr	Gly	Ser	Gly	Ser	Asp	Lys	Ile
			100					105					110		
Glu	Trp	Ala	Arg	Met	Lys	Val	Gly	Ser	Arg	Val	Gln	Asp	Thr	Lys	Tyr
		115					120					125			
Gly	Gln	His	Trp	Ile	Gly	Val	Gln	His	Gln	Gln	Val	Ser			
	130					135					140				

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 199 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met	Ala	Asp	Thr	His	Arg	Val	Asp	Arg	Thr	Asp	Arg	His	Phe	Gln	Phe
1				5					10					15	
Gln	Ser	Pro	Tyr	Glu	Gly	Gly	Arg	Gly	Gln	Gly	Gln	Tyr	Glu	Gly	Asp
			20					25					30		
Arg	Gly	Tyr	Gly	Gly	Gly	Gly	Tyr	Lys	Ser	Met	Met	Pro	Glu	Ser	Gly
		35					40					45			
Pro	Ser	Ser	Thr	Gln	Val	Leu	Ser	Leu	Leu	Ile	Gly	Val	Pro	Val	Val
	50					55					60				
Gly	Ser	Leu	Ile	Ala	Leu	Ala	Gly	Leu	Leu	Leu	Ala	Gly	Ser	Val	Ile
65				70						75					80
Gly	Leu	Met	Val	Ala	Leu	Pro	Leu	Phe	Leu	Ile	Phe	Ser	Pro	Val	Ile
				85				90						95	
Val	Pro	Ala	Gly	Leu	Thr	Ile	Gly	Leu	Ala	Met	Thr	Gly	Phe	Leu	Ala
			100				105						110		
Ser	Gly	Met	Phe	Gly	Leu	Thr	Gly	Leu	Ser	Ser	Ile	Ser	Trp	Val	Met
		115					120					125			
Asn	Tyr	Leu	Arg	Gly	Thr	Ala	Arg	Thr	Val	Pro	Glu	Gln	Leu	Glu	Tyr
	130					135					140				
Ala	Lys	Arg	Arg	Met	Ala	Asp	Ala	Val	Gly	Tyr	Ala	Gly	Gln	Lys	Gly
145					150					155					160
Lys	Glu	Met	Gly	Gln	His	Val	Gln	Asn	Lys	Ala	Gln	Asp	Val	Lys	Gln
				165					170					175	

WO 98/45461

PCT/US98/07179

-69-

Tyr Asp Ile Ser Lys Pro His Asp Thr Thr Thr Lys Gly His Glu Thr
 180 185 190

Gln Gly Gly Thr Thr Ala Ala
 195

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 199 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Met Ala Asp Thr His Arg Val Asp Arg Thr Asp Arg His Phe Gln Phe
 1 5 10 15

Gln Ser Pro Tyr Glu Gly Gly Arg Gly Gln Gly Gln Tyr Glu Gly Asp
 20 25 30

Arg Gly Tyr Gly Gly Gly Gly Tyr Lys Ser Met Met Pro Glu Ser Gly
 35 40 45

Pro Ser Ser Thr Gln Val Leu Ser Leu Leu Ile Gly Val Pro Val Val
 50 55 60

Gly Ser Leu Ile Ala Leu Ala Gly Leu Leu Ile Ala Gly Ser Val Ile
 65 70 75 80

Gly Leu Met Val Ala Leu Pro Leu Phe Leu Ile Phe Ser Pro Val Ile
 85 90 95

Val Pro Ala Ala Leu Thr Ile Gly Leu Ala Met Thr Gly Phe Leu Ala
 100 105 110

Ser Gly Met Phe Gly Leu Thr Gly Leu Ser Ser Ile Ser Trp Val Met
 115 120 125

Asn Tyr Leu Arg Gly Thr Arg Arg Thr Val Pro Glu Gln Leu Glu Tyr
 130 135 140

Ala Lys Arg Arg Met Ala Asp Ala Val Gly Tyr Ala Gly Gln Lys Gly
 145 150 155 160

Lys Glu Met Gly Gln His Val Gln Asn Lys Ala Gln Asp Val Lys Gln
 165 170 175

Tyr Asp Ile Ser Lys Pro His Asp Thr Thr Thr Lys Gly His Glu Thr
 180 185 190

WO 98/45461

PCT/US98/07179

-70-

Gln Gly Arg Thr Thr Ala Ala
195

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1267 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GAGCTCGATC ACACAAAGAA AACGTCAAAT GGATCATACT GGGCCCATT TGCAGACCAA	60
GAGAAAGTGA GAGAGAGTTG TCCTCTCGTT ATCAAGTAAC AGTAGACCAC CACTAAACCG	120
CCAATAGCTT ATAATCAAAA TAGAAAGGTC TAATAACAGA AACAAATGAA AAAGCCTTGT	180
TCCATGGACT GCCTACCCGA ATTGATTGAT TCGACTAGTT TTTCTTCTTC TTTGATTAAG	240
ACCTCCGTAA GAAAAATGGT ACTACTAAAG CCACTCGCTA CCAAAACTAA ACCATTCCAG	300
ACTGTAACTG GACCAATATT TCTAAACTGT AACCAGATCT CAAACATATA AACTAATTAA	360
GAACTATAAC CATTAACCGT AAAAATAAAT TTACTACAGT AAAAAATTAT ACTAATTTCA	420
GCTATGATGG AATTTTCAGCT CTTAAGAGTT GTGGAAATCA AGTAAACCTA AAATCCTAAT	480
AATATTCCTC ATCCTTATTT TTGTTTCACA TGCATGCTGT CCAATCTGTT ATTAGCATTT	540
GAAAGCCTAA AATTCTATAT ACAGTACAAT AAATCTAATT AATTTTCATT ACTAATAAAA	600
TGCTTCATAT ATACTCTTGT ATTTATAAAT CATCCGTTAT CGTTACTATA CCTTTATACA	660
TCATCCTACA TTCATACCTA AGCTAGCAAA GCAAACACT ACTAAAGGGTTCG TCAACGCAAG	720
TTATTTGCTA GTTGGTGCAT ACTACACACG GCTACGGCAA CATTAAGTAA CACATTAAGA	780
GGTGT TTTCT TAATGTAGTA TGGTAATTAT ATTTATTTCA AAACCTTGGAT TAGATATAAA	840
GGTACAGGTA GATGAAAAAT ATTTGGTTAG CGGGTTGAGA TTAAGCGGAT ATAGGAGGCA	900
TATATACAGC TGTGAGAAGA AGAGGGATAA ATACAAAAAG GGAAGGATGT TTTTGCCGAC	960
AGAGAAAGGT AGATTAAGTA GGCATCGAGA GGAGAGCAAT TGTAATAATGG ATGATTTGTT	1020
TGGTTTTGTA CGGTGGAGAG AAGAACGAAA AGATGATCAG GTAAAAAATG AAACCTTGGAA	1080
ATCATGCAAA GCCACACCTC TCCCTTCAAC ACAGTCTTAC GTGTCGTCTT CTCTTCACTC	1140
CATATCTCCT TTTTATTACC AAGAAATATA TGTCAATCCC ATTTATATGT ACGTTCTCTT	1200

WO 98/45461

PCT/US98/07179

-71-

AGACTTATCT CTATATACCC CCTTTTAATT TGTGTGCTCT TAGCCTTTAC TCTATAGTTT 1260
TAGATAG 1267

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GTAATACGAC TCACTATAGG GC

22

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GGGGATCCTA TACTAAACT ATAGAGTAAA GG

32

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Trp Ile Gly His Asp Ala Gly His
1 5

WO 98/45461

PCT/US98/07179

-72-

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 8 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Asn Val Gly His Asp Ala Asn His
1 5

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 8 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Val Leu Gly His Asp Cys Gly His
1 5

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 8 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Val Ile Ala His Glu Cys Gly His
1 5

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 8 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

WO 98/45461

PCT/US98/07179

-73-

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Val Ile Gly His Asp Cys Ala His
1 5

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Val Val Gly His Asp Cys Gly His
1 5

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

His Asn Ala His His
1 5

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

His Asn Tyr Leu His His
1 5

WO 98/45461

PCT/US98/07179

-74-

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

His Arg Thr His His
1 5

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

His Arg,Arg His His
1 5

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

His Asp Arg His His
1 5

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

WO 98/45461

PCT/US98/07179

-75-

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

His Asp Gln His His
1 5

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

His Asp His His His
1 5

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

His Asn His His His
1 5

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Phe Gln Ile Glu His His
1 5

WO 98/45461

PCT/US98/07179

-77-

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

His	Ile	Pro	His	His
1				5

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

His	Val	Pro	His	His
1				5

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1941 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GAGCTCGATC ACACAAAGAA AACGTCAAAT GGATCATACT GGGCCCATTT TGCAGACCAA	60
GAGAAAGTGA GAGAGAGTTG TCCTCTCGTT ATCAAGTAAC AGTAGACCAC CACTAAACCG	120
CCAATAGCTT ATAATCAAAA TAGAAAGGTC TAATAACAGA AACAAATGAA AAAGCCTTGT	180
TCCATGGACT GCCTACCCGA ATTGATTGAT TCGACTAGTT TTTCTTCTTC TTTGATTAAG	240
ACCTCCGTAA GAAAAATGGT ACTACTAAAG CCACTCGCTA CCAAACTAA ACCATTCCAG	300
ACTGTAAGT GACCAATATT TCTAAACTGT AACCAGATCT CAAACATATA AACTAATTAA	360
GAACTATAAC CATTAACCGT AAAAATAAAT TTACTACAGT AAAAAATTAT ACTAATTTCA	420
GCTATGATGG AATTTTCAGCT CTTAAGAGTT GTGGAAATCA AGTAAACCTA AAATCCTAAT	480
AATATTCTTC ATCCTTATTT TTGTTTCACA TGCATGCTGT CCAATCTGTT ATTAGCATTT	540
GAAAGCCTAA AATTCTATAT ACAGTACAAT AAATCTAATT AATTTTCATT ACTAATAAAA	600

WO 98/45461

PCT/US98/07179

- 78 -

TGCTTCATAT	ATACTCTTGT	ATTTATAAAT	CATCCGTTAT	CGTTACTATA	CCTTTATACA	660
TCATCCTACA	TTCATACCTA	AGCTAGCAAA	GCAAACACT	AAAAGGGTCG	TCAACGCAAG	720
TTATTTGCTA	GTTGGTGCAT	ACTACACACG	GCTACGGCAA	CATTAAGTAA	CACATTAAGA	780
GGTGTTTTCT	TAATGTAGTA	TGGTAATTAT	ATTTATTTCA	AAACTTGGAT	TAGATATAAA	840
GGTACAGGTA	GATGAAAAAT	ATTTGGTTAG	CGGGTTGAGA	TTAAGCGGAT	ATAGGAGGCA	900
TATATACAGC	TGTGAGAAGA	AGAGGGATAA	ATACAAAAAG	GGAAGGATGT	TTTTGCCGAC	960
AGAGAAAGGT	AGATTAAGTA	GGCATCGAGA	GGAGAGCAAT	TGTAAAATGG	ATGATTTGTT	1020
TGGTTTTGTA	CGGTGGAGAG	AAGAACGAAA	AGATGATCAG	GTAAAAAATG	AAACTTGGAA	1080
ATCATGCAAA	GCCACACCTC	TCCCTTCAAC	ACAGTCTTAC	GTGTCGTCTT	CTCTTCACTC	1140
CATATCTCCT	TTTTATTACC	AAGAAATATA	TGTCAATCCC	ATTTATATGT	ACGTTCTCTT	1200
AGACTTATCT	CTATATACCC	CCTTTTAATT	TGTGTGCTCT	TAGCCTTTAC	TCTATAGTTT	1260
TAGATAGACA	TGGCGAATGT	GGATCGTGAT	CGGCGTGTGC	ATGTAGACCG	TACTGACAAA	1320
CGTGTTTCATC	AGCCAAACTA	CGAAGATGAT	GTCGGTTTTG	GTGGCTATGG	CGGTTATGGT	1380
GCTGGTTCTG	ATTATAAGAG	TCGCGGCCCC	TCCACTAACC	AAGTATTTTT	GTGGTCTCTT	1440
TAGTTTTTCT	TGTGTTTTCC	TATGATCACG	CTCTCCAAAC	TATTTGAAGA	TTTTCTGTAA	1500
ATTCATTTTA	AACAGAAAGA	TAAATAAAAT	AGTGAAGAAC	CATAGGAATC	GTACGTTACG	1560
TTAATTATTT	CCTTTTAGTT	CTTAAGTCCT	AATTAGGATT	CCTTTAAAAG	TTGCAACAAT	1620
CTAATTGTTC	ACAAAATGAG	TAAAGTTTGA	AACAGATTTT	TATACACCAC	TTGCATATGT	1680
TTATCATGGT	GATGCATGCT	TGTTAGATAA	ACTCGATATA	ATCAATACAT	GCAGATCTTG	1740
GCACTTATAG	CAGGAGTCCA	TTGGTGGCAC	ACTGCTAACC	CTAGCTGGAC	TCACTCTAGC	1800
CGGTTCGGTG	ATCGGCTTGC	TAGTCTCCAT	ACCCCTCTTC	CTCCTCTTCA	GTCCGGTGAT	1860
AGTCCCGGCG	GCTCTCACTA	TTGGGCTTGC	TGTGACGGGA	ATCTTGGCTT	CTGGTTTGTT	1920
TGGGTTGACG	GGTCTGAGCT	C				1941

- 79 -

What is claimed is:

1. An isolated nucleic acid encoding an oleosin 5' regulatory region which directs seed-specific expression selected from the groups consisting of the nucleotide sequence set forth in SEQ ID NO:12, the nucleotide sequence set forth in SEQ ID NO:12 having an insertion, deletion, or substitution of one or more nucleotides, or a contiguous fragment of the nucleotide sequence set forth in SEQ ID NO:12.
2. An expression cassette which comprises the oleosin 5' regulatory region of Claim 1 operably linked to at least one of a nucleic acid encoding a heterologous gene or a nucleic acid encoding a sequence complementary to a native plant gene.
3. The expression cassette of Claim 2 wherein the heterologous gene is at least one of a fatty acid synthesis gene or a lipid metabolism gene.
4. The expression cassette of Claim 3 wherein the heterologous gene is selected from the group consisting of an acetyl-coA carboxylase gene, a ketoacyl synthase gene, a malonyl transacylase gene, a lipid desaturase gene, an acyl carrier protein (ACP) gene, a thioesterase gene, an acetyl transacylase gene, or an elongase gene.
5. The expression cassette of Claim 4 wherein the lipid desaturase gene is selected from the group consisting of a $\Delta 6$ -desaturase gene, a $\Delta 12$ -desaturase gene, and a $\Delta 15$ -desaturase gene.
6. An expression vector which comprises the expression cassette of any one of Claims 2-5.

- 80 -

7. A cell comprising the expression cassette of any one of Claims 2-5.

8. A cell comprising the expression vector of Claim 6.

9. The cell of Claim 7 wherein said cell is a bacterial cell or a plant cell.

10. The cell of Claim 8 wherein said cell is a bacterial cell or a plant cell.

11. A transgenic plant comprising the expression cassette of any one of Claims 2-5.

12. A transgenic plant comprising the expression vector of Claim 6.

13. A plant which has been regenerated from the plant cell of Claim 9.

14. A plant which has been regenerated from the plant cell of Claim 10.

15. The plant of Claim 12 or 13 wherein said plant is at least one of a sunflower, soybean, maize, cotton, tobacco, peanut, oil seed rape or *Arabidopsis* plant.

16. Progeny of the plant of Claim 11 or 12.

17. Seed from the plant of Claim 11 or 12.

18. A method of producing a plant with increased levels of a product of a fatty acid synthesis gene or a lipid metabolism gene which comprises:

(a) transforming a plant cell with an expression vector comprising the isolated nucleic acid of Claim 1 operably linked to at least one of an

- 81 -

isolated nucleic acid coding for a fatty acid synthesis gene or a lipid metabolism gene; and

(b) regenerating a plant with increased levels of the product of said fatty acid synthesis or said lipid metabolism gene from said plant cell.

19. A method of producing a plant with increased levels of gamma linolenic acid (GLA) content which comprises:

(a) transforming a plant cell with an expression vector comprising the isolated nucleic acid of Claim 1 operably linked to a $\Delta 6$ -desaturase gene; and

(b) regenerating a plant with increased levels of GLA from said plant cell.

20. The method of Claim 19 wherein said $\Delta 6$ -desaturase gene is at least one of a cyanobacterial $\Delta 6$ -desaturase gene or a Borage $\Delta 6$ -desaturase gene.

21. The method of any one of Claims 18-20 wherein said plant is a sunflower, soybean, maize, tobacco, cotton, peanut, oil seed rape or *Arabidopsis* plant.

22. The method of Claim 18 wherein said fatty acid synthesis gene or said lipid metabolism gene is at least one of a lipid desaturase, an acyl carrier protein (ACP) gene, a thioesterase gene an elongase gene, an acetyl transacylase gene, an acetyl-coA carboxylase gene, a ketoacyl synthase gene, or a malonyl transacylase gene.

23. A method of inducing production of at least one of gamma linolenic acid (GLA) or

- 82 -

octadecatetraeonic acid (OTA) in a plant deficient or lacking in GLA which comprises transforming said plant with an expression vector comprising an the isolated nucleic acid of Claim 1 operably linked to a $\Delta 6$ -desaturase gene and regenerating a plant with increased levels of at least one of GLA or OTA.

24. A method of decreasing production of a fatty acid synthesis or lipid metabolism gene in a plant which comprises:

(a) transforming a plant cell with an expression vector comprising the isolated nucleic acid of Claim 1 operably linked to a nucleic acid sequence complementary to a fatty acid synthesis or lipid metabolism gene; and

(b) regenerating a plant with decreased production of said fatty acid synthesis or said lipid metabolism gene.

25. A method of cosuppressing a native fatty acid synthesis or lipid metabolism gene in a plant which comprises:

(a) transforming a cell of the plant with an expression vector comprising the isolated nucleic acid of Claim 1 operably linked to a nucleic acid sequence encoding a fatty acid synthesis or lipid metabolism gene native to the plant; and

(b) regenerating a plant with decreased production of said fatty acid synthesis or said lipid metabolism gene.

2
ata tct gcc tac cct ccc aaa gag agt agt cat ttt tca tca atg gct gct caa atc aag
H A A O I K
62
aaa tac att acc tca gat gaa ctg aag aac cac gat aaa ccc gga gat cta tgg atc tgg
K Y I T S D E L K N H D K P G D L W I S
122
att caa ggg aaa gcc tat gat gtt tgg gat tgg gtc aaa gac cat cca ggt ggc agc ttt
I Q G K A Y D V S D W V K D H P G G S F
182
ccc tgg aag agt ctt gct ggt caa gag gta act gat gca ttt gtt gca ttc cat cct gcc
P L K S L A G O E V T D A F V A F H P A
242
tct aca tgg aag aat ctt gat aag ttt ttc act ggg tat tat ctt aaa gat tac tct gtt
S T W K H L D K F F T G Y Y L K D Y S V
302
tct gag gtt tct aaa gat tat agg aag ctt gtc ttt gag ttt tct aaa atg ggt ttg tat
S E V S K D Y R K L V F E F S K H G L Y
362
gac aaa aaa ggt cat att atg ttt gca act ttg tgc ttt ata gca atg ctg ttt gct atg
D K K G H I H F A T L C F I A H L F A H
422
agt gtt tat ggg gtt ttg ttt tgt gag ggt gtt ttg gta cat ttg ttt tct ggg tgt ttg
S V Y G V L F C E G V L V H L F S O C L
482
atg ggg ttt ctt tgg att cag agt ggt tgg att gga cat gat gct ggg cat tat atg gta
M G F L W I O S G W I D H A D H Y H V
542
gtg tct gat tca agg ctt aat aag ttt atg ggt att ttt gct gca aat tgt ctt tca gga
V S D S R L N K F M G I F A A H C L S G
602
ata agt att ggt tgg tgg aaa tgg aac cat aat gca cat cac att gcc tgt aat agc ctt
I S I G W W K W N H N A H H I A C N S L
662
gaa tat gac cct gat tta caa tat ata cca ttc ctt gtt gtc tct tcc aag ttt ttt ggt
E Y D P N L Q Y I P F L V V S S K F F G
722
tca ctg acc tct cat ttc tat gag aaa agg ttg act ttt gac tct tta tca aga ttc ttt
S L T S H F Y E K R L T F D S L S R F F
782
gta agt tat caa cat tgg aca ttt tac cct att atg tgt gct gct agg ctg aat atg tat
V S Y Q H W T F Y P I H C A A R L N H Y
842
gta caa tct ctg ata atg ttg ttg acc aag aga aat gtc tcc tat cga gct cag gaa ctg
V O S L I H L L T K R N V S Y R A Q E L
902
ttg gga tgc cta gtc ttc tgg att tgg tac cgg ttg ctt gtt tct tgt ttg cct aat tgg
L G C L V F S I W Y P L L V S C L P H W
962
ggt gaa aga att atg ttt gtt att gca agt tta tca gtc act gga atg caa caa gtt cag
G E R I H F V I A S L S V T G H Q O V O
1022
ttc tcc ttg aac cac ttc tct tca agt gtt tat gtt gga aag cct aaa ggg aat aat tgg
F S L N H F S S S V Y V G K P K G H H W
1082
ttt gag aaa caa acg gat ggg aca ctt gac att tct tgt cct cct tgg atg gat tgg ttt
F E K O T D G T L D I S C P P W H D W F
1142
cat ggt gga ttg caa ttc caa att gag cat cat ttg ttt ccc aag atg cct aga tgc aac
H G G L O F D I P H H L F P K H P R C N
1202
ctt agg aaa atc tgg ccc tac gtc atc gag tta tgc aag aaa cat aat ttg cct tac aat
L R K I S P Y V I E L C K K H H L P Y H
1262
tat gca tct ttc tcc aag gcc aat gaa atg aca ctg aga aca ttg agg aac aca gca ttg
Y A S F S K A H E H T L R T L R H T A L
1322
cag jct agg gat ata acc aag cgg ctg cgg aag aat ttg gta tgg gaa gct ctt cac aat
Q A R D I T K P L P K N L V W E A L H T
1382
cat ggt taa aat tac cct tag ttc atg taa taa ttt gag att atg tat ctg cta tgt ttg
H G
1442
tgt ctt gtc ttg gtt cta ctt gtt gga gtc att gca act tgt ctt tta tgg ttt att aga
1502
tgt ttt tta ata tat ttt aga ggt ttt gct ttc atc tcc att att gat gaa taa gga gtt
1562
gca tat tgt caa ttg ttg tgc tca ata tct gat att ttg gaa tgt act ttg tac caa tgt
1622
gct ttc agt tga agc tca tgt gta ctt cta tag act ttg ttt aaa tgg tta tgt cat gtt
1682
att t

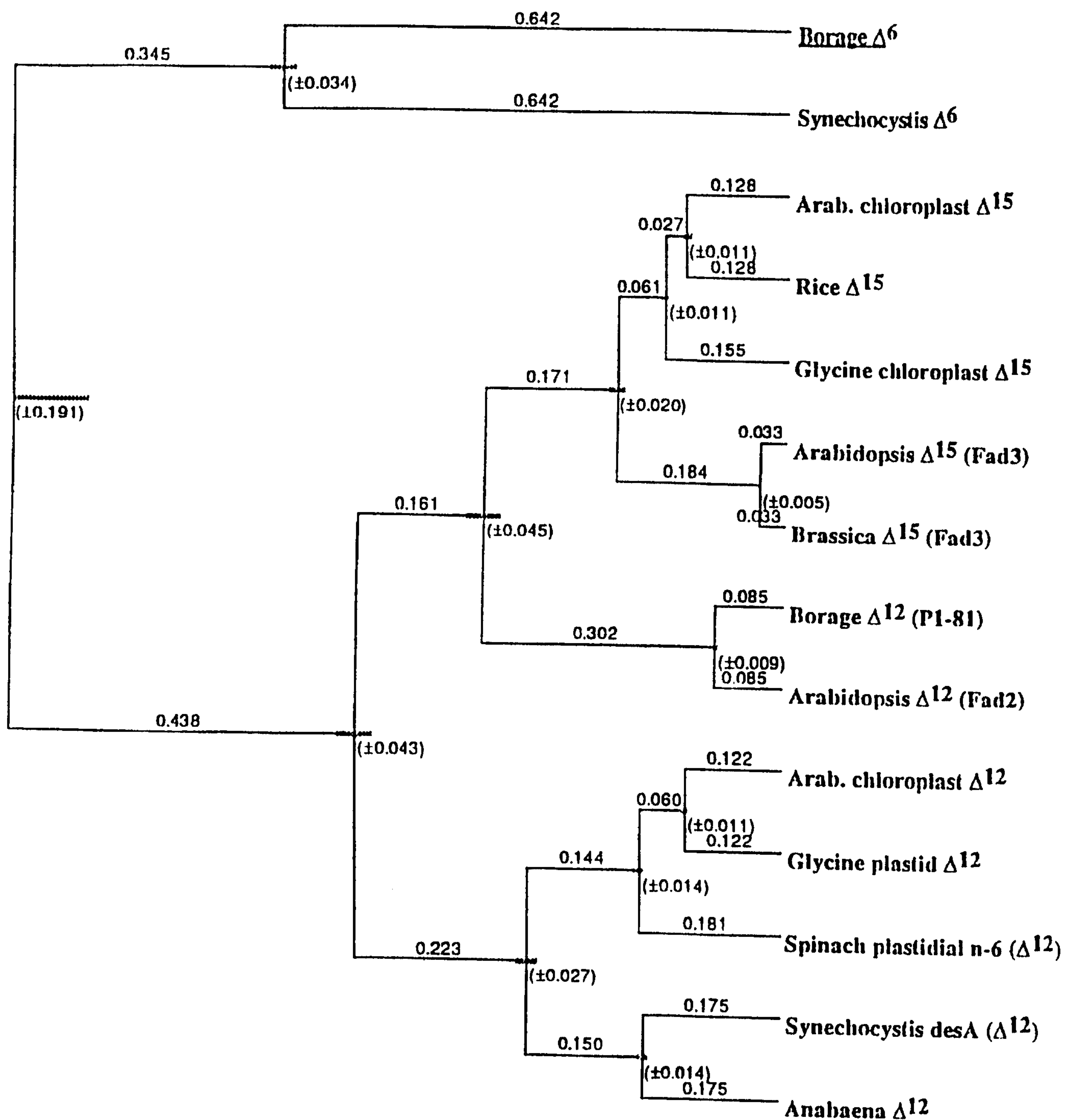


FIGURE 2

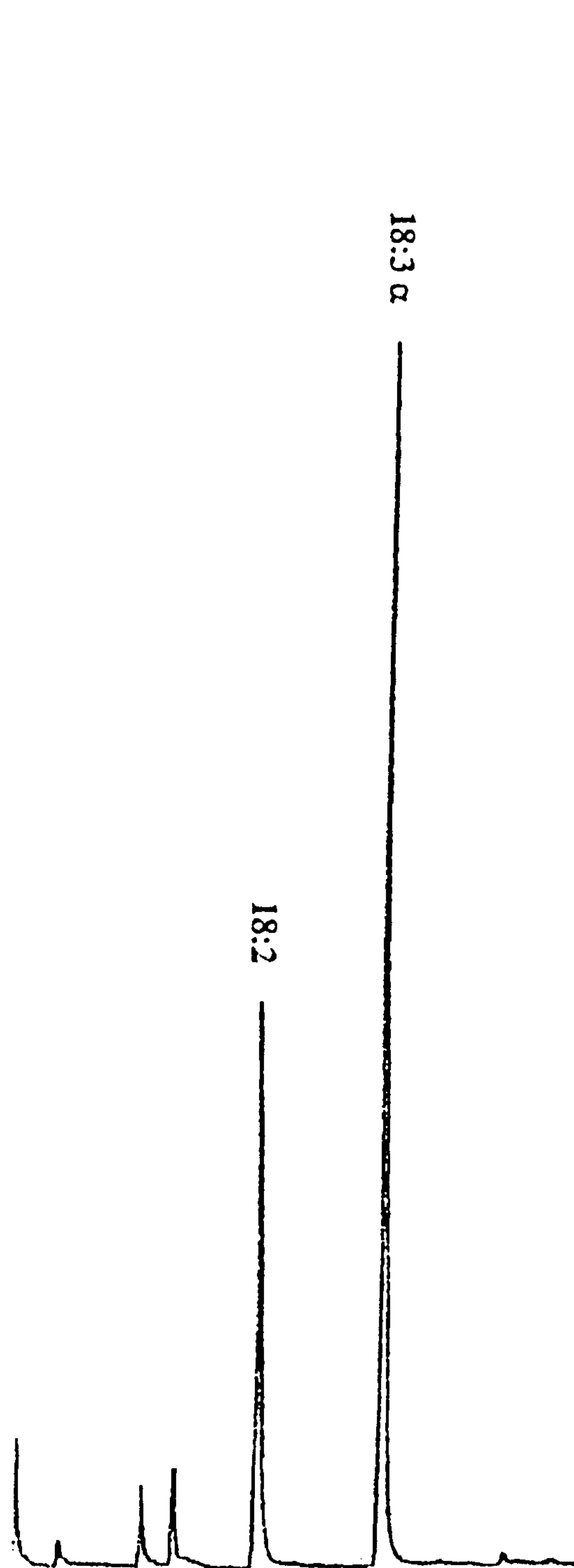


FIGURE 3A

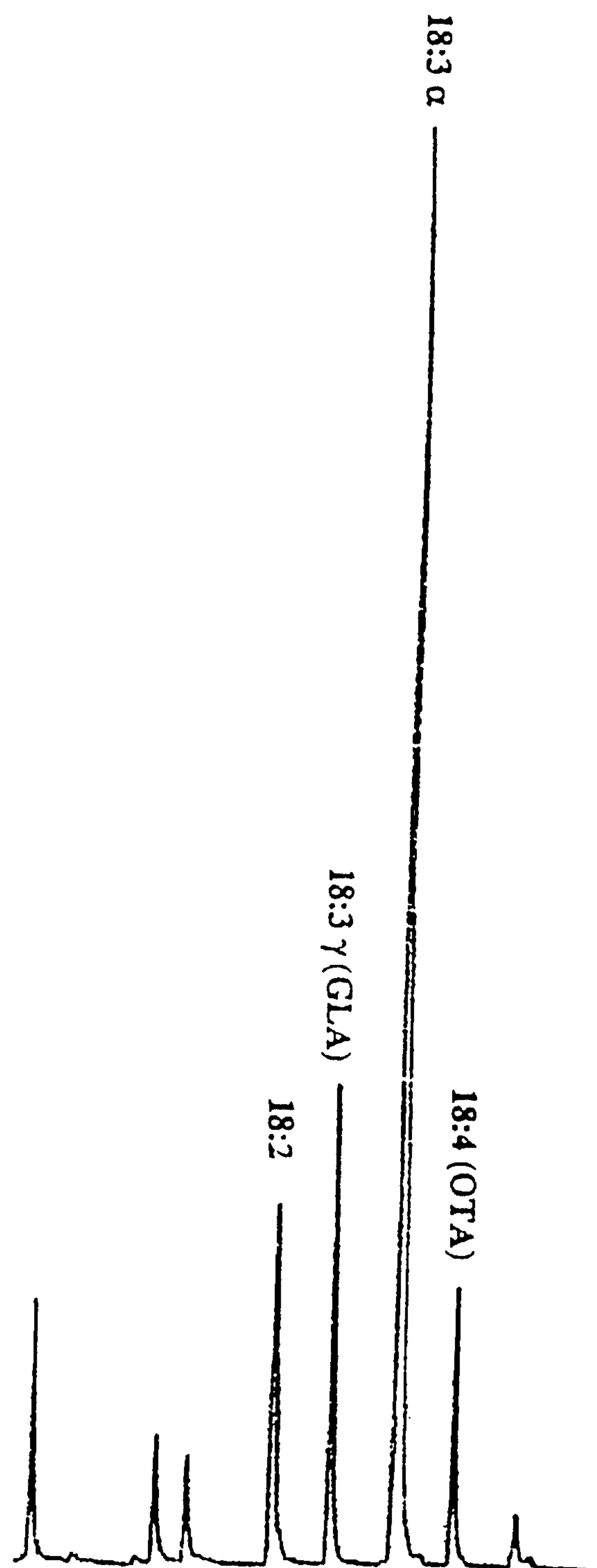


FIGURE 3B

FIGURE 4

1 tta gcc ttt act cta tag ttt tag ata gac atg gcg aat gtg gat cgt gat cgg cgt gtg
1 M A N V D R D R R V
61 cat gta gac cgt act gac aaa cgt gtt cat cag cca aac tac gaa gat gat gtc ggt ttt
11 H V D R T D K R V H Q P N Y E D D V G F
121 ggt ggc tat ggc ggt tat ggt gct ggt tct gat tat aag agt cgc ggc ccc tcc act aac
31 G G Y G G Y G A G S D Y K S R G P S T N
181 caa atc ttg gca ctt ata gca gga gtt ccc att ggt ggc aca ctg cta acc cta gct gga
51 Q I L A L I A G V P I G G T L L T L A G
241 ctc act cta gcc ggt tcg gtg atc ggc ttg cta gtc tcc ata ccc ctc ttc ctc ctc ttc
71 L T L A G S V I G L L V S I P L F L L F
301 agt ccg gtg ata gtc ccg gcg gct ctc act att ggg ctt gct gtg acg gga atc ttg gct
91 S P V I V P A A L T I G L A V T G I L A
361 tct ggt ttg ttt ggg ttg acg ggt ctg agc tcg gtc tcg tgg gtc ctc aac tac ctc cgt
111 S G L F G L T G L S S V S W V L N Y L R
421 ggg acg agt gat aca gtg cca gag caa ttg gac tac gct aaa cgg cgt atg gct gat gcg
131 G T S D T V P E Q L D Y A K R R M A D A
481 gta ggc tat gct ggt atg aag gga aaa gag atg ggt cag tat gtg cna gat aag gct cat
151 V G Y A G M K G K E M G Q Y V Q D K A H
541 gag gct cgt gag act gag ttc atg act gag acc cat gag ccg ggt aag gcc agg aga ggc
171 E A R E T E F M T E T H E P G K A R R G
601 tca taa gct aat ata aat tgc ggg agt cag ttg gaa acg cga taa atg tag ttt tac ttt
191 S *
661 tat gtc cca gtt tct ttc ctc ttt taa gaa tat ctt tgt cta tat atg tgt tcg ttc gtt
721 ttg tct tgt cca aat aaa aat cct tgt tag tga aat aag aaa tga aat aaa tat gtt ttc
781 ttt ttt gag ata acc aga aat ctc ata cta ttt tct aaa aaa aaa aaa aaa aaa

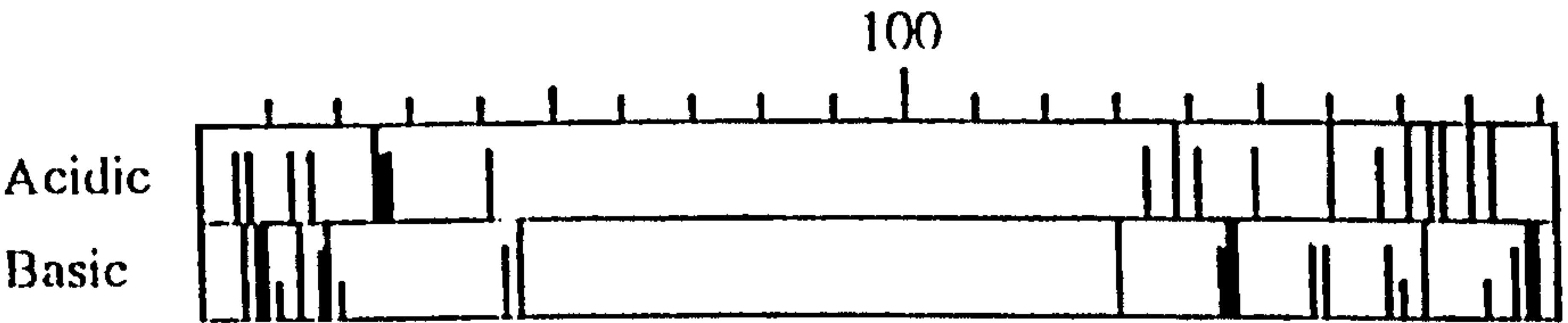


FIGURE 5

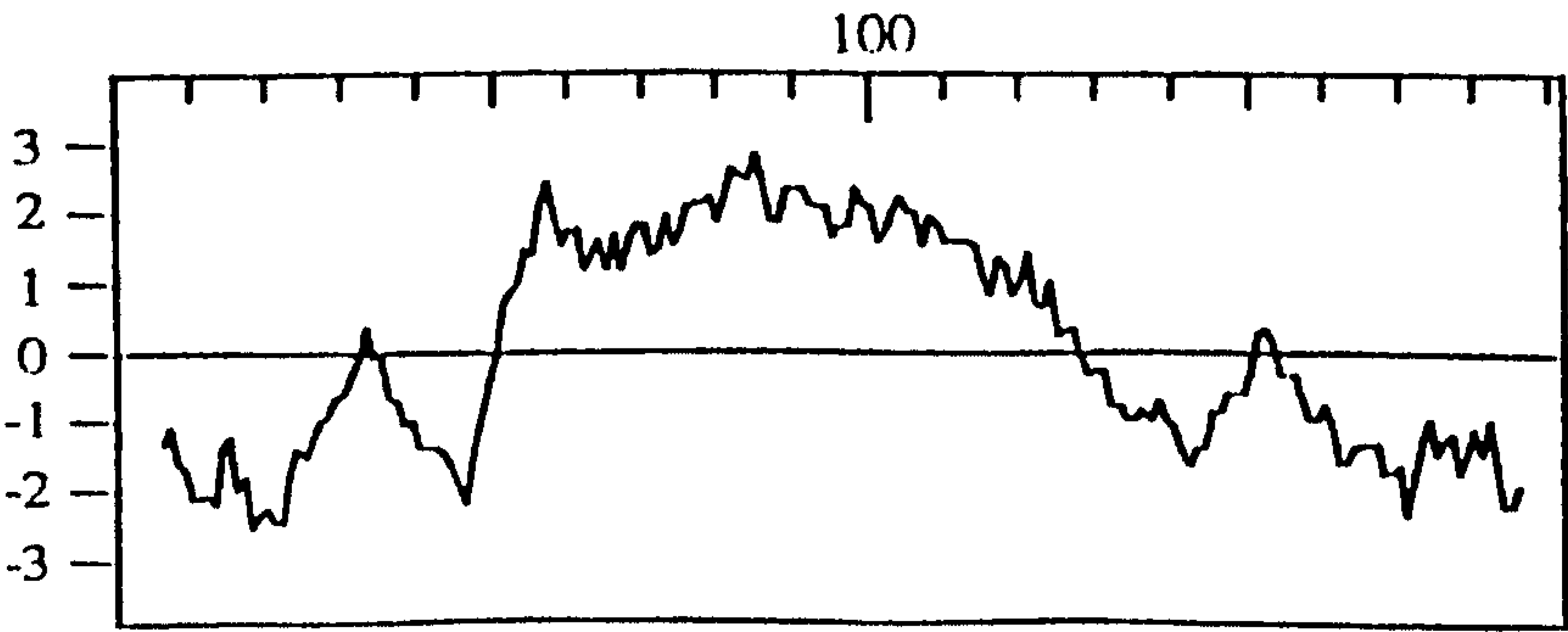


FIGURE 6

WO 98/45461

5/14

PCT/US98/07179

AtS21	MANVDRDRRV	HVDRTDKRVH	-QPNYEDDVG	---F-G--GY	GGYGAGSDYK	43
X91918	MANVDRDRRV	HVDRTDKRVH	-QPNYEDDVG	---F-G--GY	GGYAGSDYK	43
Z29859	-----	-----	-----	-----	-----	
X62352	MADTARG---	-THEDYIGND	QYPMNORDND	QYQMSGRG--	-----SDY-	37
Atol3	MADQTR----	-THEE-----	---MISRDSF	Q-----	-----HAN-	21
X91956	MADTHR----	-VDRTDRHFQ	FQSPYEGGRG	QGQYEGDRGY	GGGGYKSMMP	45
L40954	MADTHR----	-VDRTDRHFQ	FQSPYEGGRG	QGQYEGDRGY	GGGGYKSMMP	45
AtS21	SRGPSTNCIL	ALIAGVPIGG	TLITIIAGLTI	AGSVIGLIIVS	IPIFLIFSPV	93
X91918	SRGPSTNCIL	ALIAGVPIGG	TLITIIAGLTI	AGSVIGLIIVS	IPIFLIFSPV	93
Z29859	-----	-----	-----	-----	-----	
X62352	---SKSRCLA	KAATAVTAGG	SLIVLSGLTI	EGTVIALIIVA	TPILVIFSPV	84
Atol3	---PKSRCLW	KAATAVTAGG	SLIVLSGLTI	EGTVIALIIVA	TPILVIFSPV	68
X91956	ESGPSSTQVL	SLLIQVPVVG	SLIALIAGLLI	AGSVIGIIMVA	IPIFLIFSPV	95
L40954	ESGPSSTQVL	SLLIQVPVVG	SLIALIAGLLI	AGSVIGIIMVA	IPIFLIFSPV	95
AtS21	IVPAALIIIGL	AVTGIIASGL	FGLTGLSSVS	WVLNYLIGTS	DTVPEQLDVA	143
X91918	IVPAALIIIGL	AVTGIIASGL	FGLTGLSSVS	WVLNYLIGTS	DTVPEQLDVA	143
Z29859	-----	-----	FGLTGLSSVS	WVLNYLIGTS	DTVPEQLDVA	30
X62352	IVPAALIIIVAI	LIITGFISSGG	FGIAAITVFS	WLYNYATGSH	PGSDKLDVA	134
Atol3	IVPAALIIIVAI	LIITGFISSGG	FGIAAITVFS	WLYNYATGSH	PGSDKLDVA	115
X91956	IVPAALIIIGL	AMTGFIIASGM	FGLTGLSSIS	WMNYLRGTA	RTVPEQLEYA	145
L40954	IVPAALIIIGL	AMTGFIIASGM	FGLTGLSSIS	WMNYLRGTA	RTVPEQLEYA	145
AtS21	KRRMADAVGY	AGMKGKEMGQ	YVQOKAHEAR	ETEF-----	MTETHEPGKA	187
X91918	KRRMADAVGY	AGMKGKEMGQ	YVQOKAHEAR	ETEF-----	MTETHEPGKA	187
Z29859	KRRMADAVGY	AGMKGKEMGQ	YVQOKAHEAR	ETEF-----	MTETHEPGKA	74
X62352	RMKLGSKA--	QDIKIRA-DY	YGOGETGGSH	DDDTAGGQH	TT-----	173
Atol3	RMKVGSIV--	QDIK-----	YGOENIGVQH	QQ-----	VS-----	141
X91956	KRRMADAVGY	AGCKGKEMGQ	HVONKAQDVK	QYDISKPHDT	TTKGHETQGS	195
L40954	KRRMADAVGY	AGCKGKEMGQ	HVONKAQDVK	QYDISKPHDT	TTKGHETQGS	195
AtS21	RRG3 =	Z54164/Atol2				191
X91918	RRG3					191
Z29859	RRG3					78
X62352	----					173
Atol3	----					141
X91956	TTAA					199
L40954	TTAA					199

FIGURE 7

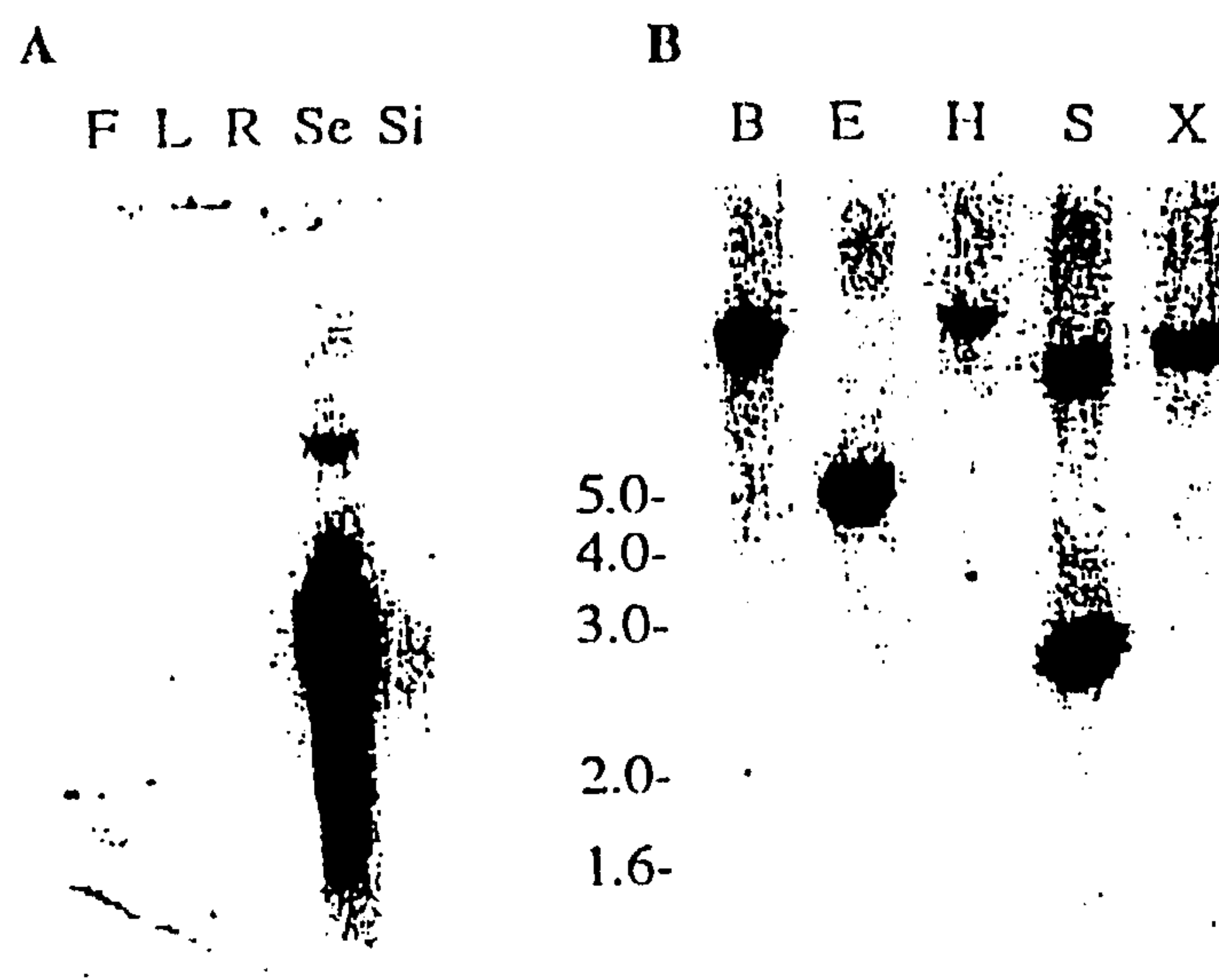


FIGURE 8A

FIGURE 8B

FIGURE 9

A

1 GAGCTCGATCACACAAAGAAACGTCAAATGGATCATACTGGGCCCATTTTGCAGACCAAGAGAAAGTGAGAGAGAGTTG
 81 TCCTCTCGTTATCAAGTAACAGTAGACCACCCTAAACCGCCAATAGCTTATAATCAAAATACAAAGGTCTAATAACAGA
 161 AACAAATGAAAAAGCCTTGTTCCATGGACTGCCCTACCCGAATTGATTGATTGACTAGTTTTCTTCTTCTTTGATTAAAG
 241 ACCTCCGTAAGAAAAATGGTACTACTAAAGCCACTCGCTACCAAACTAAACCATTCCAGACTGTAAGTGGACCAATATT
 321 TCTAAACTGTAACCAGATCTCAACATATAAACTAATTAAGAACTATAACCATTAAACCGTAAAAATAAATTTACTACAGT
 401 AAAAAATTATACTAATTTACAGCTATGATGGAATTTACAGCTCTTAAGAGTTGTGGAAATCAAGTAAACCTAAAATCCTAAT
 481 AATATTCTTCATCCTTATTTTTGTTTCACATGCATGCTGTCCAATCTGTTATTAGCATTTGAAAGCCTAAAATTCATATAT
 561 ACAGTACAATAAATCTAATTAATTTTCATTACTAATAAAATGCTTCATATATACTCTTGTATTTATAAATCATCCGTTAT
 641 CGTTACTATACCTTTATACATCATCCTACATTTCATACCTAAGCTAGCAAAGCAAACACTACTAAAAGGGTCGTCAACGCAAG
 721 TTATTTGCTAGTTGGTGCATACTACACACGGCTACGGCAACATTAAGTAACACATTAAGAGGTGTTTTCTTAATGTAGTA
 801 TGGTAATTATATTTATTTCAAACTTGGATTAGATATAAAGGTACAGGTAGATGAAAAATATTTGGTTAGCGGGTTGAGA
 881 TTAAGCGGATATAGGAGGCATATATACAGCTGTGAGAAGAAGAGGGATAAATACAAAAGGGAAGGATGTTTTGCGCGAC
 961 AGAGAAAGGTAGATTAAGTAGGCATCGAGAGGAGAGCAATTGTAAAATGGATGATTGTTTGGTTTTGTACGGTGGAGAG
 1041 AAGAACGAAAAGATGATCAGGTAAAAAATGAAACTTGGAAATCATGCAAAGCCACACCTCTCCCTTCAACACAGTCTTAC
 1121 GTGTCGTCTTCTCTTCACTCCATATCTCCTTTTTATTACCAAGAAATATATGTCAATCCCATTATATGTACGTTCTCTT
 1201 AGACTTATCTCTATATACCCCTTTTAATTTGTGTGCTCcttagcctttactctatagttttagatagacatgycgaatgt
 1281 ggatcgtgatcggcgtgtgcatgtagaccgtactgacaaacgtgttcacagccaaactacgaagatgatgtcggttttg
 1361 gtggctatggcggttatggtgctggttctgattataagagtcgcggccctccactaaccaaGTATTTTGTGGTCTCTT
 1441 TAGTTTTTCTTGTGTTTTCTATGATCACGCTCTCCAACCTATTTGAAGATTTTCTGTAAATTCATTTTAAACAGAAAGA
 1521 TAAATAAAATAGTGAAGAACCATAGGAATCGTACGTTACGTTAATTATTTCTTTTAGTTCTTAAGTCCTAATTAGGATT
 1601 CCTTTAAAGTTGCAACAATCTAATTGTTTCACAAAATGAGTAAAGTTTGAAACAGATTTTTATACACCACTTGCATATGT
 1681 TTATCATGGTGATGCATGCTTGTAGATAAACTCGATATAATCAATACATGCAGatcttggcacttatagcaggagtcc
 1761 catgtggtggcacactgctaaccctagctggactcactctagccggttcgggtgatcggttgctagtctccatacccctct
 1821 tccctcctcttcagtcgggtgatagtcggcggtctctcactattgggcttgctgtgacgggaatcttggcttctgggttg
 1901 tttgggttgacgggtctgagctc

B

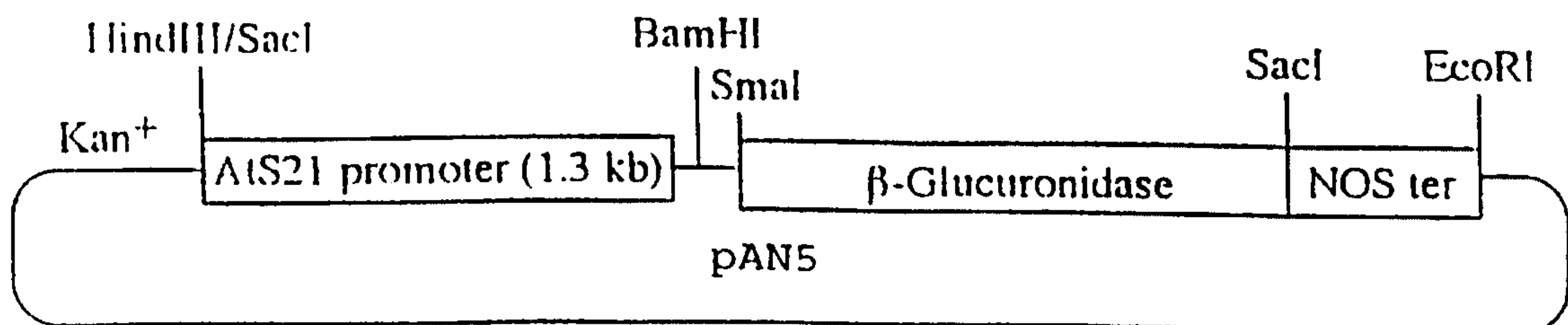


FIGURE 10

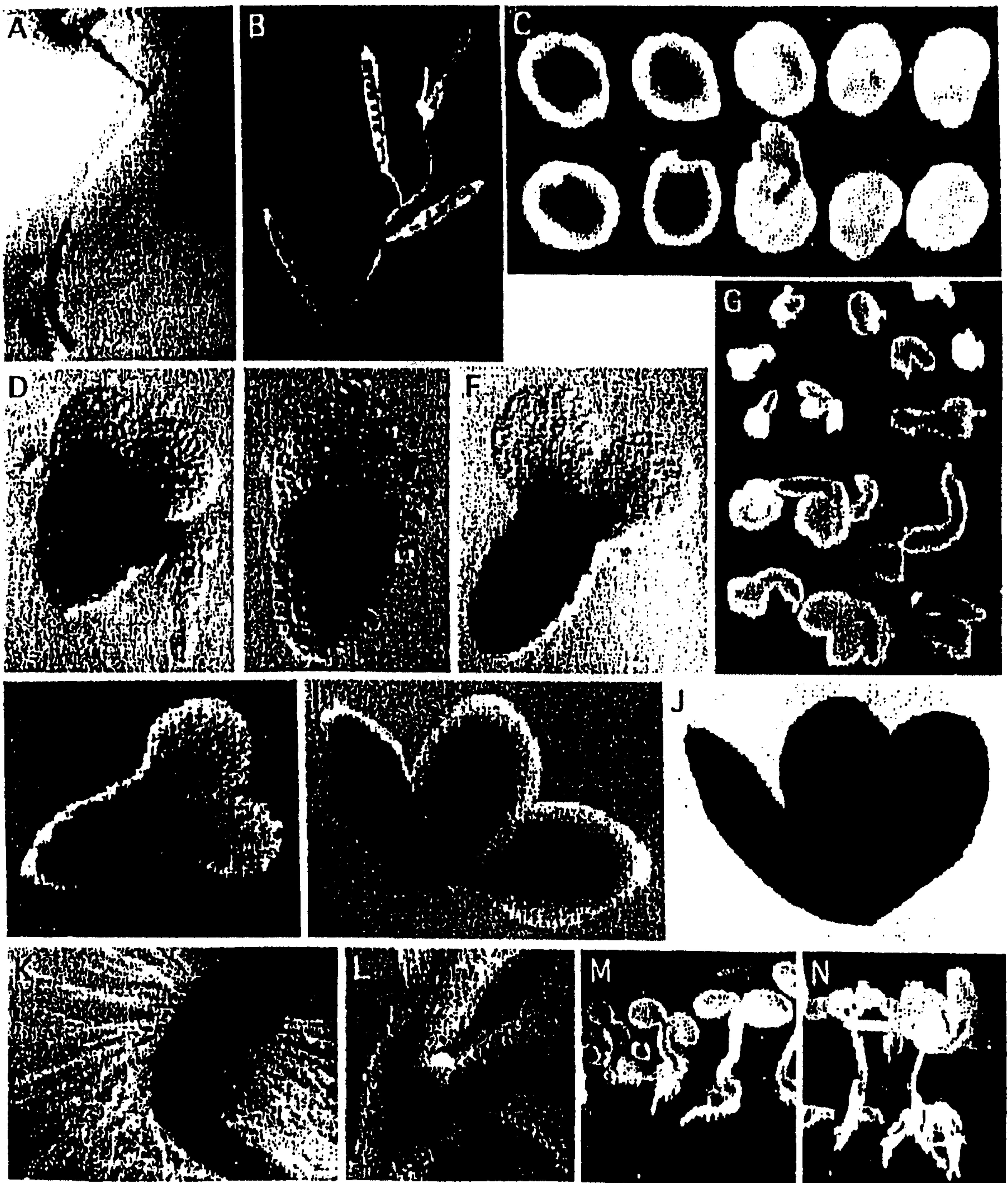


FIGURE 11A-11N

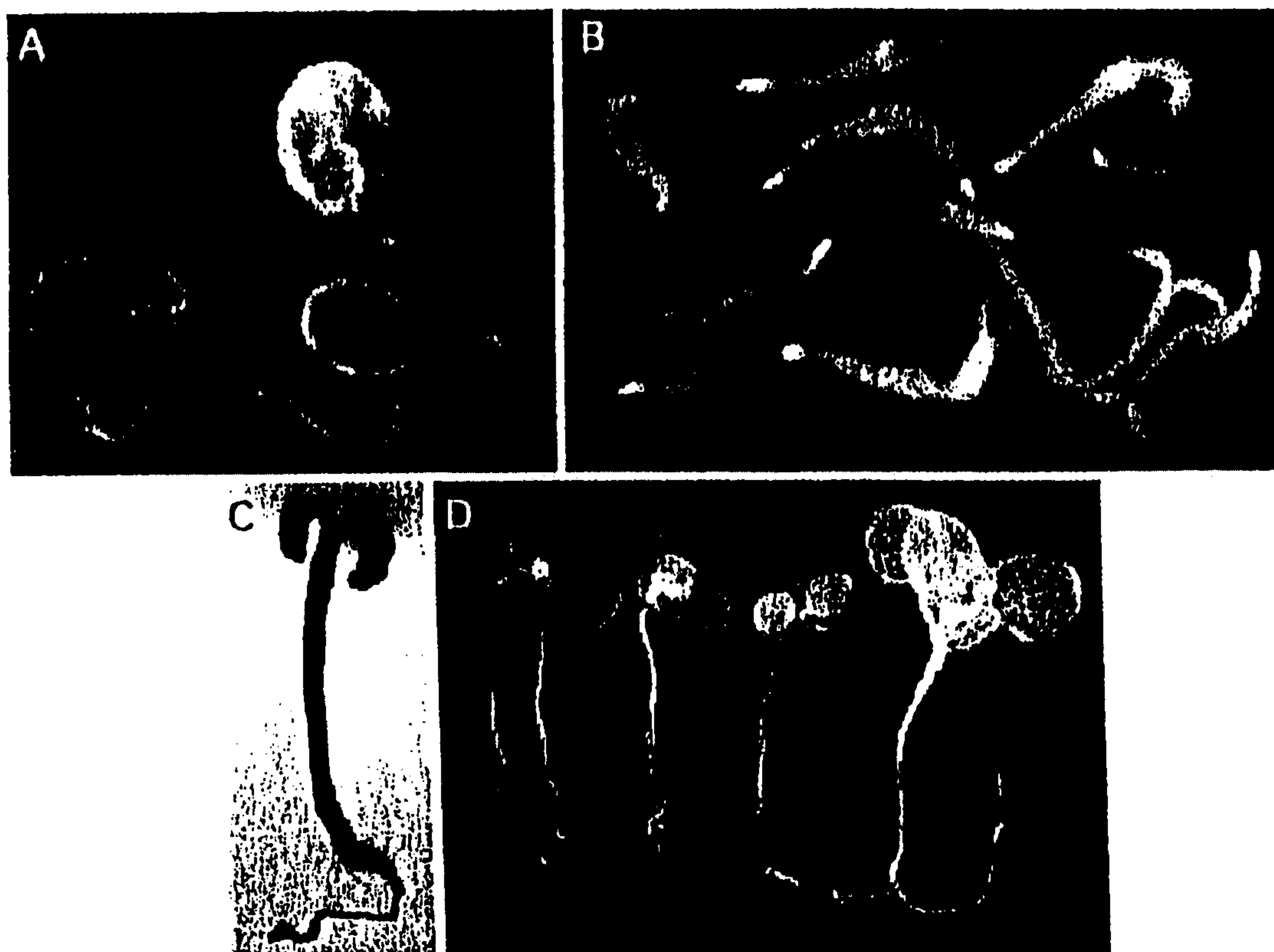


FIGURE 12A-D

WO 98/45461

10/14

PCT/US98/07179

1 2 3 4 5 6 7 8

1 2 3 4 5 6 7 8

1 2 3 4 5 6 7 8



FIGURE 13A



FIGURE 13B



FIGURE 13C

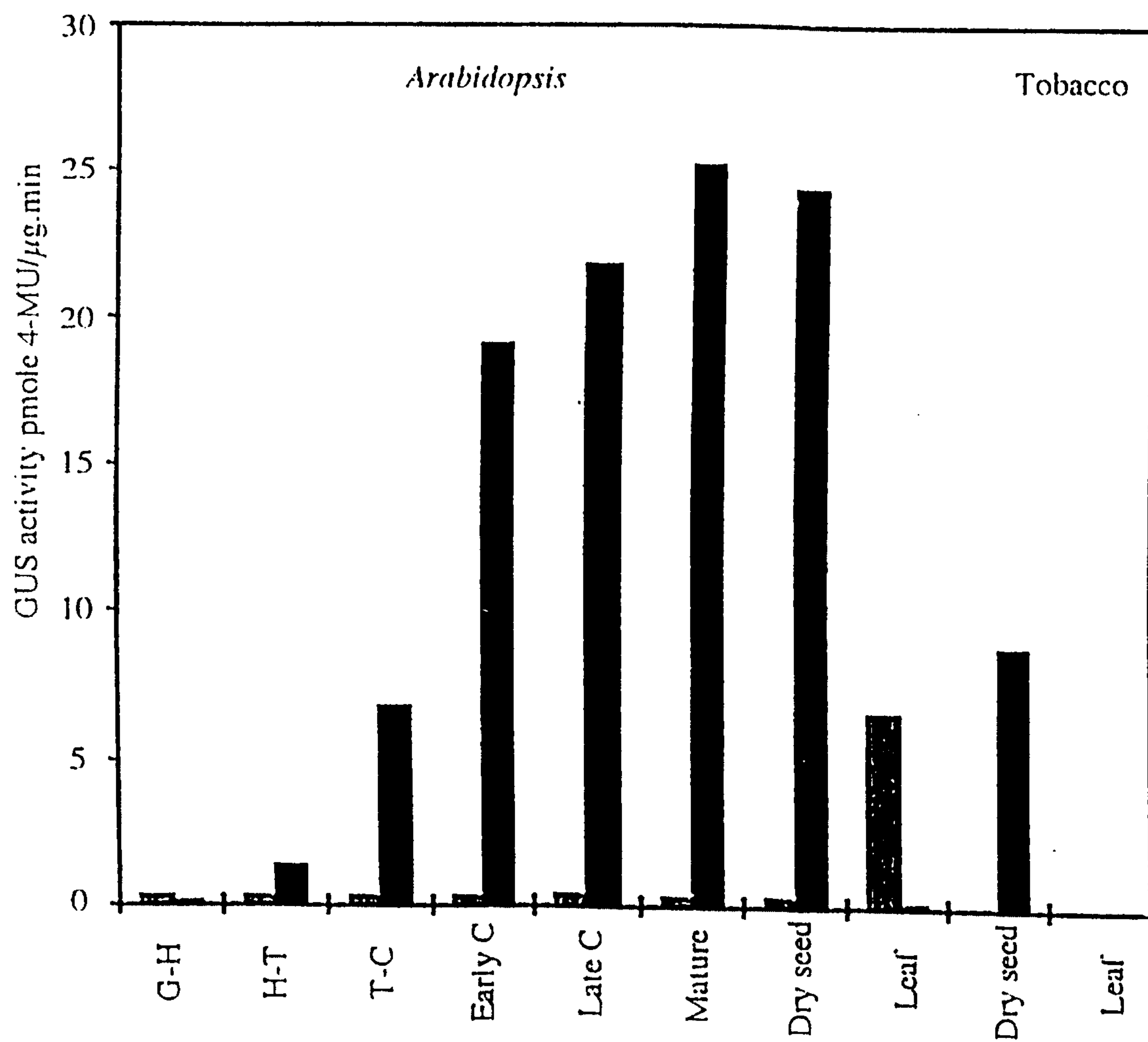


FIGURE 14

FIGURE 15B

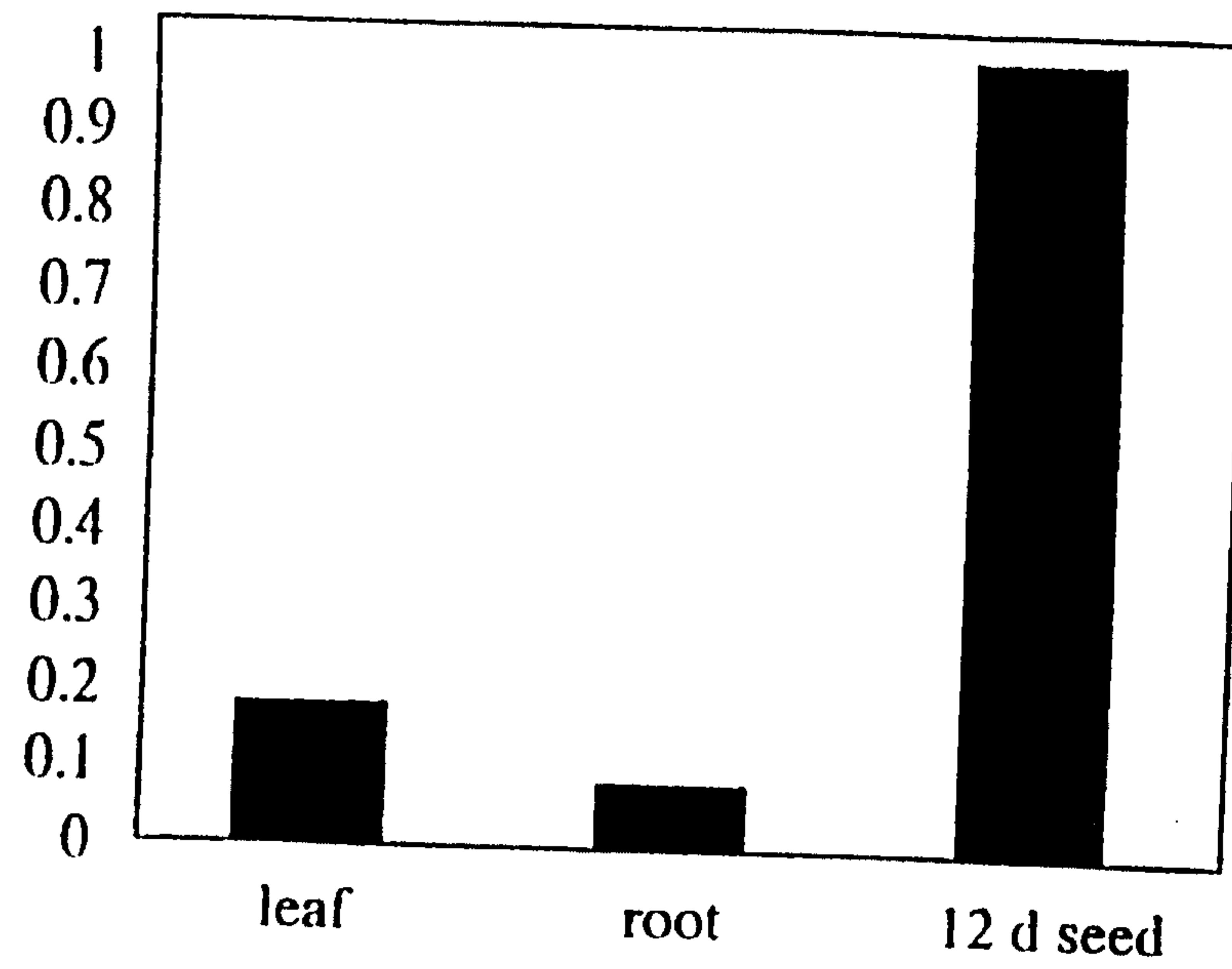
**Borage tissue****L R S**

FIGURE 15A

FIGURE 16A FIGURE 16B FIGURE 16C

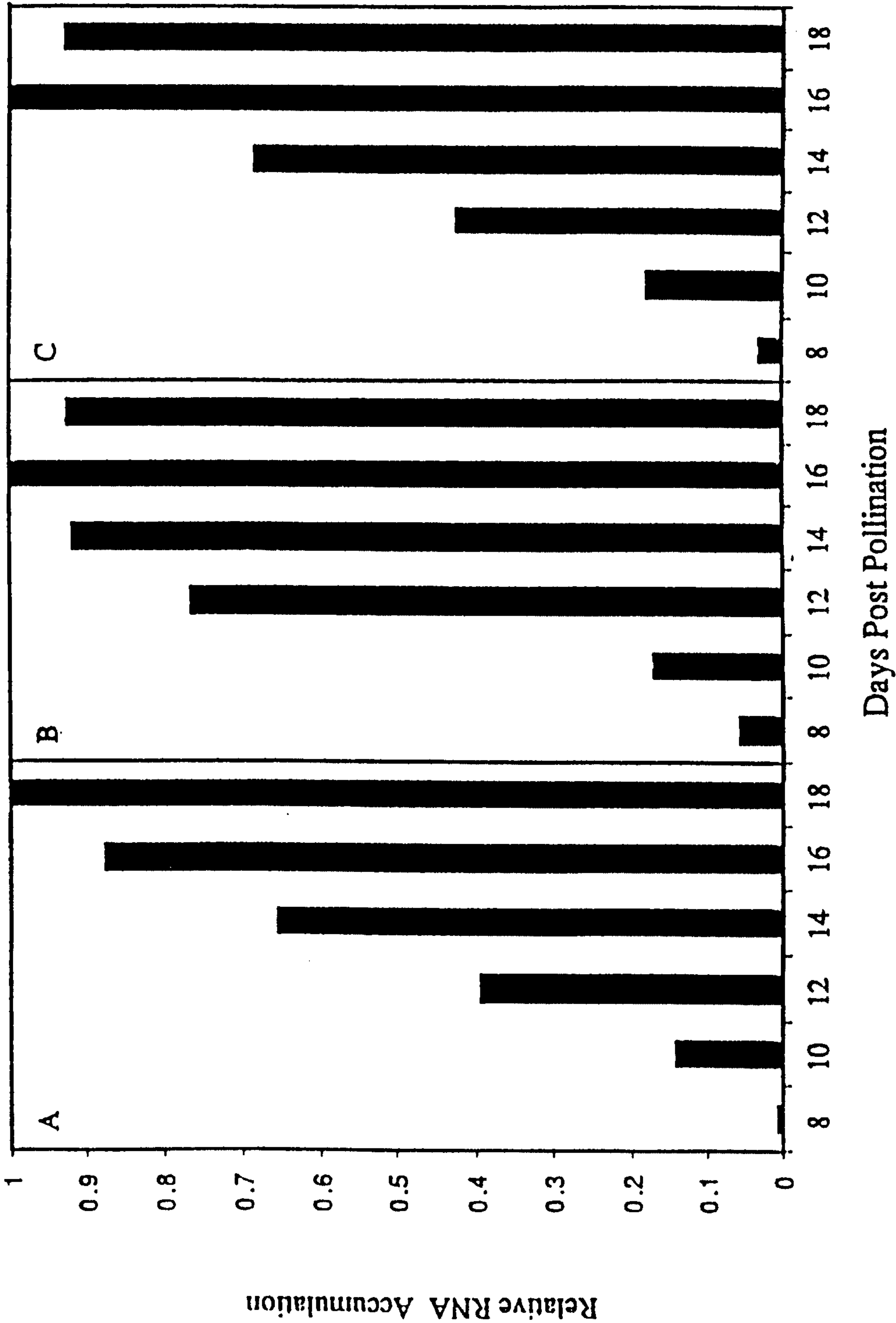


FIGURE 17

