Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2764073 A1 2011/01/20

(21) 2 764 073

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2010/06/29
(87) Date publication PCT/PCT Publication Date: 2011/01/20
(85) Entrée phase nationale/National Entry: 2011/11/30
(86) N° demande PCT/PCT Application No.: US 2010/040281
(87) N° publication PCT/PCT Publication No.: 2011/008510
(30) Priorité/Priority: 2009/06/30 (US61/221,731)

(51) Cl.Int./Int.Cl. *C12N 9/14* (2006.01), *A01H 5/00* (2006.01), *A01H 5/10* (2006.01), *C12N 15/82* (2006.01)

(71) Demandeur/Applicant:
E. I. DU PONT DE NEMOURS AND COMPANY, US

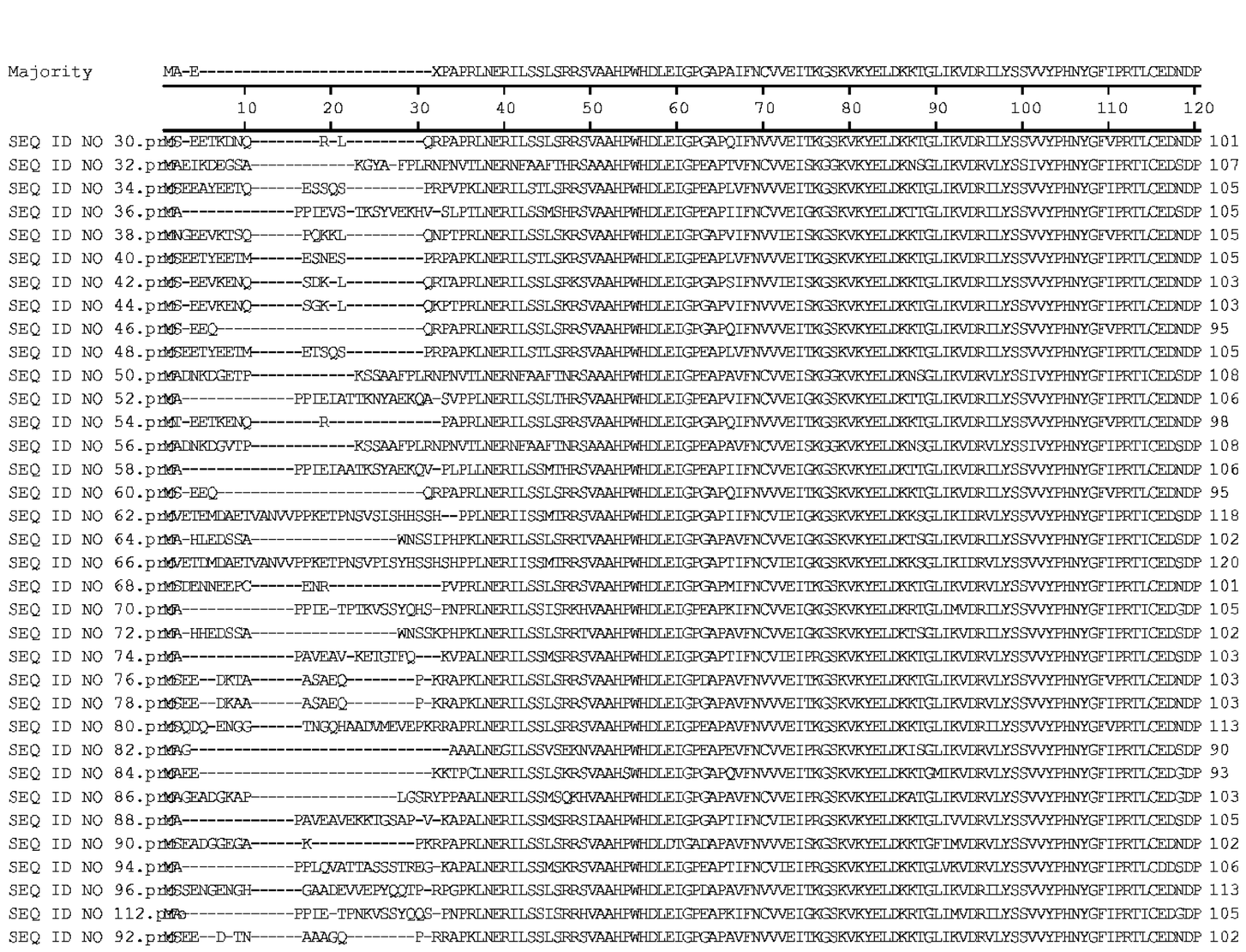
(72) Inventeurs/Inventors:
MEYER, KNUT, US;
EVERARD, JOHN D., US

(74) Agent: TORYS LLP

(54) Titre : GRAINES DE PLANTE AVEC NIVEAUX DE COMPOSE DE STOCKAGE ALTERES, PRODUITS DE CONSTRUCTION APPARENTES ET PROCEDES METTANT EN JEU DES GENES CODANT POUR LA PYROPHOSPHATASE CYTOSOLIQUE

(54) Title: PLANT SEEDS WITH ALTERED STORAGE COMPOUND LEVELS, RELATED CONSTRUCTS AND METHODS INVOLVING GENES ENCODING CYTOSOLIC PYROPHOSPHATASE

FIG. 1A



(57) Abrégé/Abstract:
This invention is in the field of plant molecular biology. More specifically, this invention pertains to isolated nucleic acid fragments encoding cytosolic pyrophosphatase proteins in plants and seeds and the use of such fragments to modulate expression of a gene encoding cytosolic pyrophosphatase activity in a transformed host cell.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 January 2011 (20.01.2011)

PCT

(10) International Publication Number
WO 2011/008510 A3

(51) International Patent Classification:
C12N 9/14 (2006.01) *A01H 5/00* (2006.01)
C12N 15/82 (2006.01) *A01H 5/10* (2006.01)

(21) International Application Number:
PCT/US2010/040281

(22) International Filing Date:
29 June 2010 (29.06.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/221,731 30 June 2009 (30.06.2009) US

(71) Applicant *(for all designated States except US)*: **E. I. DU PONT DE NEMOURS AND COMPANY** [US/US]; 1007 Market Street, Wilmington, Delaware 19898 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **MEYER, Knut** [DE/US]; 214 South Road, Lindamere, Wilmington, Delaware 19809 (US). **EVERARD, John D.** [GB/US]; 1413 Athens Road, Wilmington, Delaware 19803 (US).

(74) Agent: **EPELBAUM, Sabine, U.**; E. I. du Pont de Nemours and Company, Legal Patent Records Center, 4417 Lancaster Pike, Wilmington, Delaware 19805 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

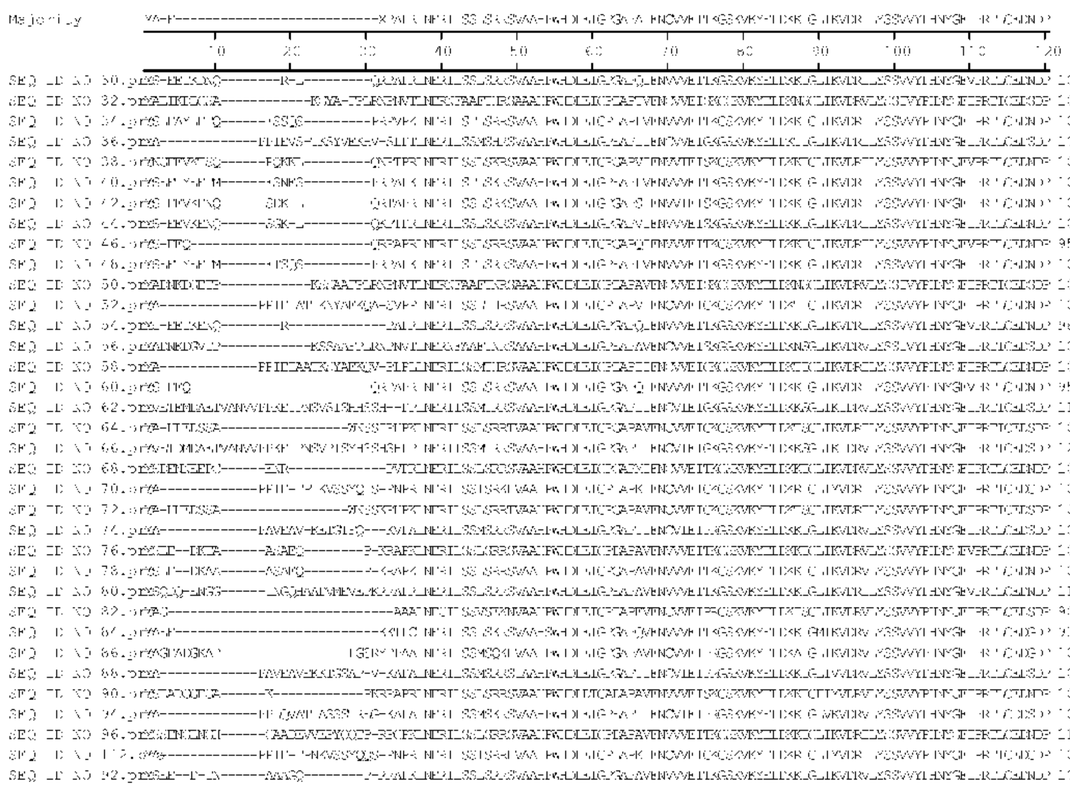
— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

[Continued on next page]

(54) Title: PLANT SEEDS WITH ALTERED STORAGE COMPOUND LEVELS, RELATED CONSTRUCTS AND METHODS INVOLVING GENES ENCODING CYTOSOLIC PYROPHOSPHATASE

FIG. 1A



(57) Abstract: This invention is in the field of plant molecular biology. More specifically, this invention pertains to isolated nucleic acid fragments encoding cytosolic pyrophosphatase proteins in plants and seeds and the use of such fragments to modulate expression of a gene encoding cytosolic pyrophosphatase activity in a transformed host cell.

WO 2011/008510 A3

WO 2011/008510 A3

— *with sequence listing part of description (Rule 5.2(a))*

(88) Date of publication of the international search report:
10 March 2011

TITLE

PLANT SEEDS WITH ALTERED STORAGE COMPOUND LEVELS, RELATED CONSTRUCTS AND METHODS INVOLVING GENES ENCODING CYTOSOLIC PYROPHOSPHATASE

This application claims the benefit of U.S. Provisional Application No. 61/221,731, filed June 30, 2009

FIELD OF THE INVENTION

This invention is in the field of plant molecular biology. More specifically, this invention pertains to isolated nucleic acid fragments encoding cytosolic pyrophosphatase proteins in plants and seeds and the use of such fragments to modulate expression of a gene encoding cytosolic pyrophosphatase activity.

BACKGROUND OF THE INVENTION

At maturity, about 40% of soybean seed dry weight is protein and 20% extractable oil. These constitute the economically valuable products of the soybean crop. Plant oils for example are the most energy-rich biomass available from plants; they have twice the energy content of carbohydrates. It also requires very little energy to extract plant oils and convert them to fuels. Of the remaining 40% of seed weight, about 10% is soluble carbohydrate. The soluble carbohydrate portion contributes little to the economic value of soybean seeds and the main component of the soluble carbohydrate fraction, raffinosaccharides, are deleterious both to processing and to the food value of soybean meal in monogastric animals (Coon et al., (1988) Proceedings Soybean Utilization Alternatives, Univ. of Minnesota, pp. 203-211).

As the pathways of storage compound biosynthesis in seeds are becoming better understood it is clear that it may be possible to modulate the size of the storage compound pools in plant cells by altering the catalytic activity of specific enzymes in the oil, starch and soluble carbohydrate biosynthetic pathways (Taiz L., et al. *Plant Physiology*; The Benjamin/Cummings Publishing Company: New York, 1991). For example, studies investigating the over-expression of LPAT and DAGAT showed that the final steps acylating the glycerol backbone exert significant control over flux to lipids in seeds. Seed oil content could also be increased in oil-seed

rape by overexpression of a yeast glycerol-3-phosphate dehydrogenase, whereas over-expression of the individual genes involved in *de novo* fatty acid synthesis in the plastid, such as acetyl-CoA carboxylase and fatty acid synthase, did not substantially alter the amount of lipids accumulated (Vigeolas H., et al. *Plant Biotechnology J*. 5, 431-441 (2007). A low-seed-oil mutant, *wrinkled 1*, has been identified in Arabidopsis. The mutation apparently causes a deficiency in the seed-specific regulation of carbohydrate metabolism (Focks, Nicole et al., Plant Physiol. (1998), 118(1), 91-101. There is a continued interest in identifying the genes that encode proteins that can modulate the synthesis of storage compounds, such as oil, protein, starch and soluble carbohydrates, in plants.

Pyrophosphatases catalyze the hydrolysis of Pyrophosphate (PPi) into two Phosphates (Pi). Pyrophosphate has been implicated in the coordination of cytosolic and plastidial carbon metabolism in the tuber of potato (Farre, Eva M. et al., Plant Physiol (2000), 132 (2), 681-688). Sonnewald, Uwe et al. (Plant J. (1992), 2(4), 571-581) generated transgenic tobacco and potato plants expressing a heterologous , bacterial pyrophosphatase gene in the cytosol in order to reduce the cytosolic pyrophosphate content. Transgenic plants showed a 3-4 fold increase in the ratio between soluble sugars and starch in source leaves compared to wild type plants.

In view of the ubiquitous nature of pyrophosphatases further investigation of their role in the regulation of storage compound content is of great interest.

SUMMARY OF THE INVENTION

In a first embodiment the present invention concerns a transgenic plant comprising a recombinant DNA construct comprising a polynucleotide operably linked to at least one regulatory element, wherein said polynucleotide encodes a polypeptide having an amino acid sequence of at least 70% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 or 112 and wherein seeds from said transgenic plant have an altered oil, protein, starch and/or soluble carbohydrate content when compared to seeds from a control plant not comprising said recombinant DNA construct.

In a second embodiment the present invention concerns transgenic seed comprising a recombinant DNA construct comprising a polynucleotide operably linked to at least one regulatory element, wherein said polynucleotide encodes a polypeptide having an amino acid sequence of at least 70% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 or 112 and wherein said transgenic seed has an altered oil, protein, starch and/or soluble carbohydrate content when compared to a control seed not comprising said recombinant DNA construct.

In a third embodiment the present invention concerns transgenic seed comprising a recombinant DNA construct comprising a polynucleotide operably linked to at least one regulatory element, wherein said polynucleotide encodes a polypeptide having an amino acid sequence of at least 70% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 or 112 and wherein said transgenic seed has an increased starch content of at least 0.5% content on a dry weight basis when compared to a control seed not comprising said recombinant DNA construct.

In a fourth embodiment the present invention concerns transgenic seed comprising:

a recombinant DNA construct comprising: (a) a polynucleotide operably linked to at least one regulatory element, wherein said polynucleotide encodes a polypeptide having an amino acid sequence of at least 70% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 or 112, or (b) a suppression DNA construct comprising at least one regulatory element operably linked to: (i) all or part of: (A) a nucleic acid sequence encoding a polypeptide having an amino acid sequence of at least 70% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: : 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 or 112 or (B) a full complement of the nucleic acid sequence of (b)(i)(A); or (ii) a region derived from all or part of a sense strand or antisense strand of a target gene of interest, said region

having a nucleic acid sequence of at least 70% sequence identity, based on the Clustal V method of alignment, when compared to said all or part of a sense strand or antisense strand from which said region is derived, and wherein said target gene of interest encodes a cytosolic Pyrophosphatase, and wherein said plant has an altered oil, protein, starch and/or soluble carbohydrate content when compared to a control plant not comprising said recombinant DNA construct.

In a fifth embodiment the invention concerns transgenic seed having an increased oil content of at least 2% on a dry-weight basis when compared to the oil content of a non-transgenic seed, wherein said transgenic seed comprises a recombinant DNA construct comprising: (a) all or part of the nucleotide sequence set forth in SEQ ID NO: 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95 or 111 ; or (b) the full-length complement of (a): wherein (a) or (b) is of sufficient length to inhibit expression of endogenous cytosolic pyrophosphatase activity in a transgenic plant and further wherein said seed has an increase in oil content of at least 2% on a dry-weight basis, as compared to seed obtained from a non-transgenic plant.

In a sixth embodiment the invention concerns transgenic seed comprising a recombinant DNA construct comprising:(a) all or part of the nucleotide sequence set forth in SEQ ID NO:29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95 or 111; or (b) the full-length complement of (a): wherein (a) or (b) is of sufficient length to inhibit expression of endogenous cytosolic pyrophosphatase activity in a transgenic plant and further wherein said seed has an increase in oil content of at least 2% on a dry-weight basis, as compared to seed obtained from a non-transgenic plant.

In a seventh embodiment the present invention concerns a method for producing transgenic seeds, the method comprising: (a) transforming a plant cell with a recombinant DNA construct comprising a polynucleotide operably linked to at least one regulatory sequence, wherein the polynucleotide encodes a polypeptide having an amino acid sequence of at least 70% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 or 112; and (b) regenerating a transgenic plant from the transformed plant cell of (a); and (c) selecting a transgenic plant that produces a

transgenic seed having an altered oil, protein, starch and/or soluble carbohydrate content, as compared to a transgenic seed obtained from a non-transgenic plant.

In an eighth embodiment the present invention concerns a method for producing transgenic seeds, the method comprising: (a) transforming a plant cell with a recombinant DNA construct comprising a polynucleotide operably linked to at least one regulatory sequence, wherein the polynucleotide encodes a polypeptide having an amino acid sequence of at least 70% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 or 112; and (b) regenerating a transgenic plant from the transformed plant cell of (a); and (c) selecting a transgenic plant that produces a transgenic seed having an increased starch content of at least 0.5% on a dry weight basis, as compared to a transgenic seed obtained from a non-transgenic plant.

In a ninth embodiment this invention concerns a method for producing transgenic seed, the method comprising: (a) transforming a plant cell with a recombinant DNA construct comprising: (i) all or part of the nucleotide sequence set forth in SEQ ID NO: 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95 or 111; or (ii) the full-length complement of (i); wherein (i) or (ii) is of sufficient length to inhibit expression of endogenous cytosolic pyrophosphatase activity in a transgenic plant; (b) regenerating a transgenic plant from the transformed plant cell of (a); and (c) selecting a transgenic plant that produces a transgenic seed having an altered oil, protein, starch and/or soluble carbohydrate content, as compared to a transgenic seed obtained from a non-transgenic plant.

In a seventh embodiment, the present invention concerns a method for producing transgenic seed, the method comprising: (a) transforming a plant cell with a recombinant DNA construct comprising: (i) all or part of the nucleotide sequence set forth in SEQ ID NO: 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95 or 111; or (ii) the full-length complement of (i); wherein (i) or (ii) is of sufficient length to inhibit expression of endogenous cytosolic pyrophosphatase activity in a transgenic plant; (b) regenerating a transgenic plant from the transformed plant cell of (a); and (c) selecting a transgenic plant that produces a transgenic seed having an increase in

oil content of at least 2% on a dry-weight basis, as compared to a transgenic seed obtained from a non-transgenic plant.

Seeds obtained from monocot and dicot plants (such as for example maize and soybean, respectively) comprising the recombinant constructs of the invention are within the scope of the present invention. Also included are seed-specific or seed-preferred promoters driving the expression of the nucleic acid sequences of the invention. Embryo or endosperm specific promoters driving the expression of the nucleic acid sequences of the invention are also included.

Furthermore the methods of the present inventions are useful for obtaining transgenic seeds from monocot plants (such as maize and rice) and dicot plants (such as soybean and canola).

Also within the scope of the invention are product(s) and/or by-product(s) obtained from the transgenic seed obtained from monocot or dicot plants, such as maize and soybean, respectively.

In another embodiment, this invention relates to a method for suppressing in a plant the level of expression of a gene encoding a polypeptide having cytosolic pyrophosphatase activity, wherein the method comprises transforming a monocot or dicot plant with any of the nucleic acid fragments of the present invention.

BRIEF DESCRIPTION OF THE DRAWING AND SEQUENCE LISTING

The invention can be more fully understood from the following detailed description and the accompanying Drawing and Sequence Listing which form a part of this application.

Figure 1A-1B shows an alignment of the amino acid sequences of cytosolic pyrophosphatase encoded by the nucleotide sequences derived from the following: *Arabidopsis thaliana* (SEQ ID NO:30, 32, 34, 36, and 38); canola (SEQ ID NO:40, 42, 44, 46, 48, 50, 52, 54, 56, 58, and 60); soybean (SEQ ID NO:62, 64, 66, 68, 70, 72, and 112); corn (SEQ ID NO:74, 76, 78, 80, and 82), and rice (SEQ ID NO:84, 86, 88, 90, 92, 94, and 96). For the consensus alignment, amino acids which are conserved among all sequences at a given position, and which are contained in at least two sequences, are indicated with an asterisk (*). Dashes are used by the program to maximize alignment of the sequences. Amino acid positions for a given SEQ ID NO are given to the left of the corresponding line of sequence. Amino acid positions for the consensus alignment are given below each section of sequence.

FIG.2 shows a chart of the percent sequence identity for each pair of amino acid sequences displayed in Figs.1A-1B.

FIG.3 corresponds to vector pHSbarENDS2.

The sequence descriptions and Sequence Listing attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §1.821-1.825.

SEQ ID NO:1 corresponds to the nucleotide sequence of vector PHSbarENDS2.

SEQ ID NO:2 corresponds to the nucleotide sequence of vector pUC9 and a polylinker.

SEQ ID NO:3 corresponds to the nucleotide sequence of vector pKR85.

SEQ ID NO:4 corresponds to the nucleotide sequence of vector pKR278.

SEQ ID NO:5 corresponds to the nucleotide sequence of vector pKR407.

SEQ ID NO:6 corresponds to the nucleotide sequence of vector pKR1468.

SEQ ID NO:7 corresponds to the nucleotide sequence of vector pKR1475.

SEQ ID NO:8 corresponds to the nucleotide sequence of vector pKR92.

SEQ ID NO:9 corresponds to the nucleotide sequence of vector pKR1478.

SEQ ID NO:10 corresponds to SAIFF and genomic DNA of lo15571,

SEQ ID NO:11 corresponds to the forward primer PPA1.

SEQ ID NO:12 corresponds to the reverse primer for PPA1.

SEQ ID NO:13 corresponds to the nucleotide sequence of vector pENTR comprising PPA1.

SEQ ID NO:14 corresponds to the nucleotide sequence of vector pKR1478 –PPA1.

SEQ ID NO:15 corresponds to the nucleotide sequence of PKR1482.

SEQ ID NO:16 corresponds to the AthLcc In forward primer.

SEQ ID NO;17 corresponds to the AthLcc In reverse primer.

SEQ ID NO:18 corresponds to the PCR product with the laccase intron.

SEQ ID NO:19 corresponds to the nucleotide sequence of PSM1318.

SEQ ID NO:20 corresponds to the nucleotide sequence of pMBL18 ATTR12 INT.

SEQ ID NO:21 corresponds to the nucleotide sequence of PMS1789.

SEQ ID NO:22 corresponds to the nucleotide sequence of pMBL18 ATTR12 INT ATTR21.

SEQ ID NO:23 corresponds to the nucleotide sequence of vector pKR1480.

SEQ ID NO:24 corresponds to the PPA1 UTR forward primer.

SEQ ID NO:25 corresponds to the PPA1 UTR reverse primer.

SEQ ID NO:26 corresponds to the nucleotide sequence of pENTR containing the PPA1 3'UTR.

SEQ ID NO:27 corresponds to the nucleotide sequence of pKR1482 containing the PPA1 3'UTR.

SEQ ID NO:28 corresponds to the nucleotide sequence of pKR1482 containing the ORF of PPA1.

Table 1 lists the polypeptides that are described herein, the designation of the clones that comprise the nucleic acid fragments encoding polypeptides representing all or a substantial portion of these polypeptides, and the corresponding identifier (SEQ ID NO:) as used in the attached Sequence Listing. Table 1 also identifies the cDNA clones as individual ESTs ("EST"), the sequences of the entire cDNA inserts comprising the indicated cDNA clones ("FIS"), contigs assembled from two or more ESTs ("Contig"), contigs assembled from an FIS and one or more ESTs ("Contig*"), or sequences encoding the entire or functional protein derived from an FIS, a contig, an EST and PCR, or an FIS and PCR ("CGS").

TABLE 1

Cytosolic Pyrophosphatase Proteins

Protein (Plant Source)	Clone Designation	Status	SEQ ID NO: (Nucleotide)	SEQ ID NO: (Amino Acid)
Pyrophosphatase (PPA1) (Arabidopsis)	At1g01050	CGS	29	30
Pyrophosphatase (PPA2) (Arabidopsis)	At2g18230	CGS	31	32
Pyrophosphatase (PPA3) (Arabidopsis)	At2g46860	CGS	33	34
Pyrophosphatase (PPA4) (Arabidopsis)	At3g53620	CGS	35	36
Pyrophosphatase (PPA5) (Arabidopsis)	At4g01480	CGS	37	38

Pyrophosphatase (Canola)	TC23077	CGS	39	40
Pyrophosphatase (Canola)	TC20341	CGS	41	42
Pyrophosphatase (Canola)	TC16648	CGS	43	44
Pyrophosphatase (Canola)	TC20135	CGS	45	46
Pyrophosphatase (Canola)	TC23373	CGS	47	48
Pyrophosphatase (Canola)	DY22345.1	CGS	49	50
Pyrophosphatase (Canola)	TC34086	CGS	51	52
Pyrophosphatase (Canola)	TC22517	CGS	53	54
Pyrophosphatase (Canola)	TC56550	CGS	55	56
Pyrophosphatase (Canola)	TC26534	CGS	57	58
Pyrophosphatase (Canola)	TC16649	CGS	59	60
Pyrophosphatase (Soybean)	Glyma19g35710	CGS	61	62
Pyrophosphatase (Soybean)	Glyma01g37790	CGS	63	64
Pyrophosphatase (Soybean)	Glyma03g33000	CGS	65	66
Pyrophosphatase (Soybean)	Glyma07g05390	CGS	67	68
Pyrophosphatase (Soybean)	Glyma10g05130	CGS	69	70
Pyrophosphatase	Glyma11g07530	CGS	71	72

(Soybean)				
Pyrophosphatase (Soybean)	Glyma13g19500	CGS	111	112
Pyrophosphatase (Corn)	PCO593895	CGS	73	74
Pyrophosphatase (Corn)	PCO598466	CGS	75	76
Pyrophosphatase (Corn)	PCO640614	CGS	77	78
Pyrophosphatase (Corn)	PCO640979	CGS	79	80
Pyrophosphatase (Corn)	PCO650999	CGS	81	82
Pyrophosphatase (Rice)	LOC_Os10g26600.1	CGS	83	84
Pyrophosphatase (Rice)	LOC_OS02g47600.1	CGS	85	86
Pyrophosphatase (Rice)	LOC_Os05g02310.1	CGS	87	88
Pyrophosphatase (Rice)	LOC_Os01g64670.1	CGS	89	90
Pyrophosphatase (Rice)	LOC_Os04g59040.1	CGS	91	92
Pyrophosphatase (Rice)	LOC_Os01g74350.1	CGS	93	94
Pyrophosphatase (Rice)	LOC_Os05gg36260.1	CGS	95	96

SEQ ID NO:97 is the nucleic acid sequence of the linker described in Example 15.

SEQ ID NO:98 is the nucleic acid sequence of vector pKS133 described in Example 16.

SEQ ID NO:99 corresponds to synthetic complementary region of pKS106 and pKS124.

SEQ ID NO:100 corresponds to a synthetic complementary region of pKS133.

SEQ ID NO:101 corresponds to a synthetic PCR primer.

SEQ ID NO:102 corresponds to a synthetic PCR primer.

SEQ ID NO:103 corresponds to a synthetic PCR primer (SA5).

SEQ ID NO:104 corresponds to a synthetic PCR primer (SA7).

SEQ ID NO:105 corresponds to a synthetic PCR primer (SA6).

SEQ ID NO:106 is the nucleic acid sequence of vector pKS420.

SEQ ID NO:107 corresponds to a synthetic PCR primer (SA8).

SEQ ID NO:108 corresponds to a synthetic PCR primer (SA10).

SEQ ID NO:109 corresponds to a synthetic PCR primer (SA9).

SEQ ID NO:110 is the nucleic acid sequence of vector pKS421.

SEQ ID NO:111 is the nucleic acid sequence of a soybean pyrophosphatase

homolog (see also Table 1).

SEQ ID NO:112 is the amino acid sequence encoded by SEQ ID NO:111 (see also Table 1).

SEQ ID NO:113 corresponds to a synthetic PCR primer (SA11).

SEQ ID NO:114 corresponds to a synthetic PCR primer (SA13).

SEQ ID NO:115 corresponds to a synthetic PCR primer (SA12).

SEQ ID NO:116 is the nucleic acid sequence of vector pKS422.

SEQ ID NO:117 corresponds to a synthetic PCR primer (SA236).

SEQ ID NO:118 corresponds to a synthetic PCR primer (SA237).

SEQ ID NO:119 is the nucleic acid sequence of vector pKR1478- Glyma11g07530.

SEQ ID NO:120 corresponds to a synthetic PCR primer (SA242).

SEQ ID NO:121 corresponds to a synthetic PCR primer (SA243).

SEQ ID NO:122 is the nucleic acid sequence of vector pKR1478- PCO640614.

SEQ ID NO:123 corresponds to a synthetic PCR primer (SA245).

SEQ ID NO:124 corresponds to a synthetic PCR primer (SA246).

SEQ ID NO:125 is the nucleic acid sequence of vector pKR1478- PCO650999.

The Sequence Listing contains the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IUBMB standards described in *Nucleic Acids Res. 13*:3021-3030 (1985) and in the *Biochemical J. 219 (No. 2)*:345-373 (1984) which are herein incorporated by reference. The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

DETAILED DESCRIPTION OF THE INVENTION

All patents, patent applications, and publications cited throughout the application are hereby incorporated by reference in their entirety.

As used herein and in the appended claims, the singular forms "a", "an", and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, reference to "a plant" includes a plurality of such plants, reference to "a cell" includes one or more cells and equivalents thereof known to those skilled in the art, and so forth.

In the context of this disclosure a number of terms and abbreviations are used. The following definitions are provided.

“Open reading frame” is abbreviated ORF.

"Polymerase chain reaction" is abbreviated PCR.

"Triacylglycerols" are abbreviated TAGs.

"Co-enzyme A" is abbreviated CoA.

"Pyrophosphatase" is abbreviated PPiase.

The term "fatty acids" refers to long chain aliphatic acids (alkanoic acids) of varying chain length, from about C_{12} to C_{22} (although both longer and shorter chain-length acids are known). The predominant chain lengths are between C_{16} and C_{22}. The structure of a fatty acid is represented by a simple notation system of "X:Y", where X is the total number of carbon (C) atoms in the particular fatty acid and Y is the number of double bonds.

Generally, fatty acids are classified as saturated or unsaturated. The term "saturated fatty acids" refers to those fatty acids that have no "double bonds" between their carbon backbone. In contrast, "unsaturated fatty acids" have "double bonds" along their carbon backbones (which are most commonly in the *cis*-configuration). "Monounsaturated fatty acids" have only one "double bond" along the carbon backbone (e.g., usually between the 9^{th} and 10^{th} carbon atom as for palmitoleic acid (16:1) and oleic acid (18:1)), while "polyunsaturated fatty acids" (or "PUFAs") have at least two double bonds along the carbon backbone (e.g., between the 9^{th} and 10^{th}, and 12^{th} and 13^{th} carbon atoms for linoleic acid (18:2); and between the 9^{th} and 10^{th}, 12^{th} and 13^{th}, and 15^{th} and 16^{th} for α-linolenic acid (18:3)).

The terms "triacylglycerol", "oil" and "TAGs" refer to neutral lipids composed of three fatty acyl residues esterified to a glycerol molecule (and such terms will be used interchangeably throughout the present disclosure herein). Such oils can contain long chain PUFAs, as well as shorter saturated and unsaturated fatty acids and longer chain saturated fatty acids. Thus, "oil biosynthesis" generically refers to the synthesis of TAGs in the cell.

The term "modulation" or "alteration" in the context of the present invention refers to increases or decreases of PPiase expression, protein level or enzyme activity, as well as to an increase or decrease in the storage compound levels, such as oil, protein, starch or soluble carbohydrates.

The term "plant" includes reference to whole plants, plant parts or organs (e.g., leaves, stems, roots, etc.), plant cells, seeds and progeny of same. Plant cell, as used herein includes, without limitation, cells obtained from or found in the

following: seeds, suspension cultures, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen and microspores. Plant cells can also be understood to include modified cells, such as protoplasts, obtained from the aforementioned tissues. The class of plants which can be used in the methods of the invention is generally as broad as the class of higher plants amenable to transformation techniques, including both monocotyledonous and dicotyledonous plants.

Examples of monocots include, but are not limited to (corn) maize, wheat, rice, sorghum, millet, barley, palm, lily, *Alstroemeria*, rye, and oat.

Examples of dicots include, but are not limited to, soybean, rape, sunflower, canola, grape, guayule, columbine, cotton, tobacco, peas, beans, flax, safflower, and alfalfa.

Plant tissue includes differentiated and undifferentiated tissues or plants, including but not limited to, roots, stems, shoots, leaves, pollen, seeds, tumor tissue, and various forms of cells and culture such as single cells, protoplasm, embryos, and callus tissue.

The term "plant organ" refers to plant tissue or group of tissues that constitute a morphologically and functionally distinct part of a plant.

The term "genome" refers to the following: 1. The entire complement of genetic material (genes and non-coding sequences) is present in each cell of an organism, or virus or organelle. 2. A complete set of chromosomes inherited as a (haploid) unit from one parent. The term "stably integrated" refers to the transfer of a nucleic acid fragment into the genome of a host organism or cell resulting in genetically stable inheritance.

The terms “polynucleotide”, “polynucleotide sequence”, “nucleic acid”, nucleic acid sequence”, and “nucleic acid fragment” are used interchangeably herein. These terms encompass nucleotide sequences and the like. A polynucleotide may be a polymer of RNA or DNA that is single- or double-stranded, that optionally contains synthetic, non-natural or altered nucleotide bases. A polynucleotide in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA, synthetic DNA, or mixtures thereof.

The term "isolated" refers to materials, such as “isolated nucleic acid fragments” and/or “isolated polypeptides”, which are substantially free or otherwise

removed from components that normally accompany or interact with the materials in a naturally occurring environment. Isolated polynucleotides may be purified from a host cell in which they naturally occur. Conventional nucleic acid purification methods known to skilled artisans may be used to obtain isolated polynucleotides. The term also embraces recombinant polynucleotides and chemically synthesized polynucleotides.

The term "isolated nucleic acid fragment" is used interchangeably with "isolated polynucleotide" and is a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid fragment in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA. Nucleotides (usually found in their 5'-monophosphate form) are referred to by their single letter designation as follows: "A" for adenylate or deoxyadenylate (for RNA or DNA, respectively), "C" for cytidylate or deoxycytidylate, "G" for guanylate or deoxyguanylate, "U" for uridylate, "T" for deoxythymidylate, "R" for purines (A or G), "Y" for pyrimidines (C or T), "K" for G or T, "H" for A or C or T, "I" for inosine, and "N" for any nucleotide.

The terms "subfragment that is functionally equivalent" and "functionally equivalent subfragment" are used interchangeably herein. These terms refer to a portion or subsequence of an isolated nucleic acid fragment in which the ability to alter gene expression or produce a certain phenotype is retained whether or not the fragment or subfragment encodes an active enzyme. For example, the fragment or subfragment can be used in the design of recombinant DNA constructs to produce the desired phenotype in a transformed plant. Recombinant DNA constructs can be designed for use in co-suppression or antisense by linking a nucleic acid fragment or subfragment thereof, whether or not it encodes an active enzyme, in the appropriate orientation relative to a plant promoter sequence.

"Cosuppression" refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar native genes (U.S. Patent No. 5,231,020). Cosuppression technology constitutes the subject matter of U.S. Patent No. 5,231,020, which issued to Jorgensen et al. on July 27, 1999. The phenomenon observed by Napoli et al. in petunia was referred to as "cosuppression" since expression of both the endogenous gene and the introduced

transgene were suppressed (for reviews see Vaucheret et al., *Plant J.* 16:651-659 (1998); and Gura, *Nature* 404:804-808 (2000)).

Co-suppression constructs in plants previously have been designed by focusing on overexpression of a nucleic acid sequence having homology to an endogenous mRNA, in the sense orientation, which results in the reduction of all RNA having homology to the overexpressed sequence (see Vaucheret et al. (1998) *Plant J 16*:651-659; and Gura (2000) *Nature 404*:804-808). The overall efficiency of this phenomenon is low, and the extent of the RNA reduction is widely variable. Recent work has described the use of "hairpin" structures that incorporate all, or part, of an mRNA encoding sequence in a complementary orientation that results in a potential "stem-loop" structure for the expressed RNA (PCT Publication WO 99/53050 published on October 21, 1999). This increases the frequency of co-suppression in the recovered transgenic plants. Another variation describes the use of plant viral sequences to direct the suppression, or "silencing", of proximal mRNA encoding sequences (PCT Publication WO 98/36083 published on August 20, 1998). Both of these co-suppressing phenomena have not been elucidated mechanistically, although recent genetic evidence has begun to unravel this complex situation (Elmayan et al. (1998) *Plant Cell 10*:1747-1757).

In addition to cosuppression, antisense technology has also been used to block the function of specific genes in cells. Antisense RNA is complementary to the normally expressed RNA, and presumably inhibits gene expression by interacting with the normal RNA strand. The mechanisms by which the expression of a specific gene are inhibited by either antisense or sense RNA are on their way to being understood. However, the frequencies of obtaining the desired phenotype in a transgenic plant may vary with the design of the construct, the gene, the strength and specificity of its promoter, the method of transformation and the complexity of transgene insertion events (Baulcombe, *Curr. Biol.* 12(3):R82-84 (2002); Tang et al., *Genes Dev.* 17(1):49-63 (2003); Yu et al., *Plant Cell. Rep.* 22(3):167-174 (2003)). Cosuppression and antisense inhibition are also referred to as "gene silencing", "post-transcriptional gene silencing" (PTGS), RNA interference or RNAi. See for example U.S. Patent No. 6,506,559.

MicroRNAs (miRNA) are small regulatory RNSs that control gene expression. miRNAs bind to regions of target RNAs and inhibit their translation and, thus,

interfere with production of the polypeptide encoded by the target RNA. miRNAs can be designed to be complementary to any region of the target sequence RNA including the 3' untranslated region, coding region, etc. miRNAs are processed from highly structured RNA precursors that are processed by the action of a ribonuclease III termed DICER. While the exact mechanism of action of miRNAs is unknown, it appears that they function to regulate expression of the target gene. See, e.g., U.S. Patent Publication No. 2004/0268441 Al which was published on December 30, 2004.

The term "expression", as used herein, refers to the production of a functional end-product, be it mRNA or translation of mRNA into a polypeptide. "Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. "Co-suppression" refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Patent No. 5,231,020).

"Overexpression" refers to the production of a functional end-product in transgenic organisms that exceeds levels of production when compared to expression of that functional end-product in a normal, wild type or non-transformed organism.

"Stable transformation" refers to the transfer of a nucleic acid fragment into a genome of a host organism, including both nuclear and organellar genomes, resulting in genetically stable inheritance. In contrast, "transient transformation" refers to the transfer of a nucleic acid fragment into the nucleus, or DNA-containing organelle, of a host organism resulting in gene expression without integration or stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" organisms. The preferred method of cell transformation of rice, corn and other monocots is using particle-accelerated or "gene gun" transformation technology (Klein et al. (1987) *Nature (London) 327*:70-73; U.S. Patent No. 4,945,050), or an *Agrobacterium*-mediated method (Ishida Y. et al. (1996) *Nature Biotech. 14*:745-750). The term "transformation" as used herein refers to both stable transformation and transient transformation.

"Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein.

As stated herein, “suppression” refers to the reduction of the level of enzyme activity or protein functionality detectable in a transgenic plant when compared to the level of enzyme activity or protein functionality detectable in a plant with the native enzyme or protein. The level of enzyme activity in a plant with the native enzyme is referred to herein as “wild type” activity. The level of protein functionality in a plant with the native protein is referred to herein as “wild type” functionality. The term “suppression” includes lower, reduce, decline, decrease, inhibit, eliminate and prevent. This reduction may be due to the decrease in translation of the native mRNA into an active enzyme or functional protein. It may also be due to the transcription of the native DNA into decreased amounts of mRNA and/or to rapid degradation of the native mRNA. The term “native enzyme” refers to an enzyme that is produced naturally in the desired cell.

“Gene silencing,” as used herein, is a general term that refers to decreasing mRNA levels as compared to wild-type plants, does not specify mechanism and is inclusive, and not limited to, anti-sense, cosuppression, viral-suppression, hairpin suppression and stem-loop suppression.

The terms “homology”, “homologous”, “substantially similar” and “corresponding substantially” are used interchangeably herein. They refer to nucleic acid fragments wherein changes in one or more nucleotide bases does not affect the ability of the nucleic acid fragment to mediate gene expression or produce a certain phenotype. These terms also refer to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotides that do not substantially alter the functional properties of the resulting nucleic acid fragment relative to the initial, unmodified fragment. For example, alterations in a nucleic acid fragment which result in the production of a chemically equivalent amino acid at a given site, but do not affect the functional properties of the encoded polypeptide, are well known in the art. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a functionally equivalent product. Nucleotide changes

that result in alteration of the N-terminal and C-terminal portions of the polypeptide molecule would also not be expected to alter the activity of the polypeptide. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products. It is therefore understood, as those skilled in the art will appreciate, that the invention encompasses more than the specific exemplary sequences.

Moreover, the skilled artisan recognizes that substantially similar nucleic acid sequences encompassed by this invention are also defined by their ability to hybridize, under moderately stringent conditions (for example, 1 X SSC, 0.1% SDS, 60 °C) with the sequences exemplified herein, or to any portion of the nucleotide sequences reported herein and which are functionally equivalent to the gene or the promoter of the invention. Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms. Post-hybridization washes determine stringency conditions. One set of preferred conditions involves a series of washes starting with 6X SSC, 0.5% SDS at room temperature for 15 min, then repeated with 2X SSC, 0.5% SDS at 45 °C for 30 min, and then repeated twice with 0.2X SSC, 0.5% SDS at 50 °C for 30 min. A more preferred set of stringent conditions involves the use of higher temperatures in which the washes are identical to those above except for the temperature of the final two 30 min washes in 0.2X SSC, 0.5% SDS was increased to 60 °C. Another preferred set of highly stringent conditions involves the use of two final washes in 0.1X SSC, 0.1% SDS at 65 °C.

With respect to the degree of substantial similarity between the target (endogenous) mRNA and the RNA region in the construct having homology to the target mRNA, such sequences should be at least 25 nucleotides in length, preferably at least 50 nucleotides in length, more preferably at least 100 nucleotides in length, again more preferably at least 200 nucleotides in length, and most preferably at least 300 nucleotides in length; and should be at least 80% identical, preferably at least 85% identical, more preferably at least 90% identical, and most preferably at least 95% identical.

Substantially similar nucleic acid fragments of the instant invention may also be characterized by the percent identity of the amino acid sequences that they

encode to the amino acid sequences disclosed herein, as determined by algorithms commonly employed by those skilled in this art. Suitable nucleic acid fragments (isolated polynucleotides of the present invention) encode polypeptides that are at least 70% identical, preferably at least 80% identical to the amino acid sequences reported herein. Preferred nucleic acid fragments encode amino acid sequences that are at least 85% identical to the amino acid sequences reported herein. More preferred nucleic acid fragments encode amino acid sequences that are at least 90% identical to the amino acid sequences reported herein. Most preferred are nucleic acid fragments that encode amino acid sequences that are at least 95% identical to the amino acid sequences reported herein.

It is well understood by one skilled in the art that many levels of sequence identity are useful in identifying related polypeptide sequences. Useful examples of percent identities are 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%, or any integer percentage from 55% to 100%.

Sequence alignments and percent similarity calculations may be determined using a variety of comparison methods designed to detect homologous sequences including, but not limited to, the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Unless stated otherwise, multiple alignment of the sequences provided herein were performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS. 5*:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments and calculation of percent identity of protein sequences using the Clustal method are KTUPLE=1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. For nucleic acids these parameters are KTUPLE=2, GAP PENALTY=5, WINDOW=4 and DIAGONALS SAVED=4. After alignment of the sequences, using the Clustal V program, it is possible to obtain a “percent identity” by viewing the “sequence distances” table on the same program.

Unless otherwise stated, “BLAST” sequence identity/similarity values provided herein refer to the value obtained using the BLAST 2.0 suite of programs using default parameters (Altschul et al., *Nucleic Acids Res.* 25:3389-3402 (1997)). Software for performing BLAST analyses is publicly available, e.g., through the National Center for Biotechnology Information. This algorithm involves first

identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M = 5, N = -4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)).

"Sequence identity" or "identity" in the context of nucleic acid or polypeptide sequences refers to the nucleic acid bases or amino acid residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window.

Thus, "Percentage of sequence identity" refers to the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of

matched positions by the total number of positions in the window of comparison and multiplying the results by 100 to yield the percentage of sequence identity. Useful examples of percent sequence identities include, but are not limited to, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%, or any integer percentage from 55% to 100%. These identities can be determined using any of the programs described herein.

Sequence alignments and percent identity or similarity calculations may be determined using a variety of comparison methods designed to detect homologous sequences including, but not limited to, the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences are performed using the Clustal V method of alignment (Higgins, D.G. and Sharp, P.M. (1989) *Comput. Appl. Biosci. 5*:151-153; Higgins, D.G. et al. (1992) *Comput. Appl. Biosci. 8*:189-191) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments and calculation of percent identity of protein sequences using the Clustal method are KTUPLE=1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. For nucleic acids these parameters are KTUPLE=2, GAP PENALTY=5, WINDOW=4 and DIAGONALS SAVED=4.

It is well understood by one skilled in the art that many levels of sequence identity are useful in identifying polypeptides, from other plant species, wherein such polypeptides have the same or similar function or activity. Useful examples of percent identities include, but are not limited to, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%, or any integer percentage from 55% to 100%. Indeed, any integer amino acid identity from 50%-100% may be useful in describing the present invention. Also, of interest is any full or partial complement of this isolated nucleotide fragment.

The term “recombinant” means, for example, that a nucleic acid sequence is made by an artificial combination of two otherwise separated segments of sequence, e.g., by chemical synthesis or by the manipulation of isolated nucleic acids by genetic engineering techniques.

As used herein, “contig” refers to a nucleotide sequence that is assembled from two or more constituent nucleotide sequences that share common or overlapping regions of sequence homology. For example, the nucleotide

sequences of two or more nucleic acid fragments can be compared and aligned in order to identify common or overlapping sequences. Where common or overlapping sequences exist between two or more nucleic acid fragments, the sequences (and thus their corresponding nucleic acid fragments) can be assembled into a single contiguous nucleotide sequence.

"Codon degeneracy" refers to divergence in the genetic code permitting variation of the nucleotide sequence without affecting the amino acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic acid fragment comprising a nucleotide sequence that encodes all or a substantial portion of the amino acid sequences set forth herein. The skilled artisan is well aware of the "codon-bias" exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a nucleic acid fragment for improved expression in a host cell, it is desirable to design the nucleic acid fragment such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

The terms "synthetic nucleic acid" or "synthetic genes" refer to nucleic acid molecules assembled either in whole or in part from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form larger nucleic acid fragments which may then be enzymatically assembled to construct the entire desired nucleic acid fragment. "Chemically synthesized", as related to a nucleic acid fragment, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of nucleic acid fragments may be accomplished using well established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the nucleic acid fragments can be tailored for optimal gene expression based on optimization of the nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

"Gene" refers to a nucleic acid fragment that is capable of directing expression a specific protein or functional RNA.

"Native gene" refers to a gene as found in nature with its own regulatory sequences.

"Chimeric gene" or "recombinant DNA construct" are used interchangeably herein, and refers to any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature, or to an isolated native gene optionally modified and reintroduced into a host cell.

A chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. In one embodiment, a regulatory region and a coding sequence region are assembled from two different sources. In another embodiment, a regulatory region and a coding sequence region are derived from the same source but arranged in a manner different than that found in nature. In another embodiment, the coding sequence region is assembled from at least two different sources. In another embodiment, the coding region is assembled from the same source but in a manner not found in nature.

The term "endogenous gene" refers to a native gene in its natural location in the genome of an organism.

The term "foreign gene" refers to a gene not normally found in the host organism that is introduced into the host organism by gene transfer.

The term "transgene" refers to a gene that has been introduced into a host cell by a transformation procedure. Transgenes may become physically inserted into a genome of the host cell (*e.g.*, through recombination) or may be maintained outside of a genome of the host cell (*e.g.*, on an extrachromasomal array).

An "allele" is one of several alternative forms of a gene occupying a given locus on a chromosome. When the alleles present at a given locus on a pair of homologous chromosomes in a diploid plant are the same that plant is homozygous at that locus. If the alleles present at a given locus on a pair of homologous chromosomes in a diploid plant differ that plant is heterozygous at that locus. If a transgene is present on one of a pair of homologous chromosomes in a diploid plant that plant is hemizygous at that locus.

The term "coding sequence" refers to a DNA fragment that codes for a polypeptide having a specific amino acid sequence, or a structural RNA. The

boundaries of a protein coding sequence are generally determined by a ribosome binding site (prokaryotes) or by an ATG start codon (eukaryotes) located at the 5' end of the mRNA and a transcription terminator sequence located just downstream of the open reading frame at the 3' end of the mRNA. A coding sequence can include, but is not limited to, DNA, cDNA, and recombinant nucleic acid sequences.

"Mature" protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or pro-peptides present in the primary translation product have been removed. "Precursor" protein refers to the primary product of translation of mRNA; i.e., with pre- and pro-peptides still present. Pre- and pro-peptides may be and are not limited to intracellular localization signals.

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from post-transcriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA (mRNA)" refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a DNA that is complementary to and synthesized from an mRNA template using the enzyme reverse transcriptase. The cDNA can be single-stranded or converted into the double-stranded form using the Klenow fragment of DNA polymerase I. "Sense" RNA refers to RNA transcript that includes the mRNA and can be translated into protein within a cell or *in vitro*. "Antisense RNA" refers to an RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target isolated nucleic acid fragment (U.S. Patent No. 5,107,065). The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. "Functional RNA" refers to antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes. The terms "complement" and "reverse complement" are used interchangeably herein with respect to mRNA transcripts, and are meant to define the antisense RNA of the message.

The term "endogenous RNA" refers to any RNA which is encoded by any nucleic acid sequence present in the genome of the host prior to transformation with

the recombinant construct of the present invention, whether naturally-occurring or non-naturally occurring, i.e., introduced by recombinant means, mutagenesis, etc.

The term “non-naturally occurring” means artificial, not consistent with what is normally found in nature.

“Messenger RNA (mRNA)” refers to the RNA that is without introns and that can be translated into protein by the cell.

“cDNA” refers to a DNA that is complementary to and synthesized from a mRNA template using the enzyme reverse transcriptase. The cDNA can be single-stranded or converted into the double-stranded form using the Klenow fragment of DNA polymerase I.

“Sense” RNA refers to RNA transcript that includes the mRNA and can be translated into protein within a cell or *in vitro*.

“Antisense RNA” refers to an RNA transcript that is complementary to all or part of a target primary transcript or mRNA, and that blocks the expression of a target gene (U.S. Patent No. 5,107,065). The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5’ non-coding sequence, 3’ non-coding sequence, introns, or the coding sequence.

“Functional RNA” refers to antisense RNA, ribozyme RNA, or other RNA that may not be translated, yet has an effect on cellular processes. The terms “complement” and “reverse complement” are used interchangeably herein with respect to mRNA transcripts, and are meant to define the antisense RNA of the message.

The term “recombinant DNA construct” refers to a DNA construct assembled from nucleic acid fragments obtained from different sources. The types and origins of the nucleic acid fragments may be very diverse.

A “recombinant expression construct” contains a nucleic acid fragment operably linked to at least one regulatory element that is capable of effecting expression of the nucleic acid fragment. The recombinant expression construct may also affect expression of a homologous sequence in a host cell.

In one embodiment the choice of recombinant expression construct is dependent upon the method that will be used to transform host cells. The skilled artisan is well aware of the genetic elements that must be present on the recombinant expression construct in order to successfully transform, select and

propagate host cells. The skilled artisan will also recognize that different independent transformation events may be screened to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by, but is not limited to, Southern analysis of DNA, Northern analysis of mRNA expression, Western analysis of protein expression, or phenotypic analysis.

The term “operably linked” refers to the association of nucleic acid fragments on a single nucleic acid fragment so that the function of one is regulated by the other. For example, a promoter is operably linked with a coding sequence when it is capable of regulating the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in a sense or antisense orientation. In another example, the complementary RNA regions of the invention can be operably linked, either directly or indirectly, 5’ to the target mRNA, or 3’ to the target mRNA, or within the target mRNA, or a first complementary region is 5’ and its complement is 3’ to the target mRNA.

“Regulatory sequences” refer to nucleotides located upstream (5’ non-coding sequences), within, or downstream (3’ non-coding sequences) of a coding sequence, and which may influence the transcription, RNA processing, stability, or translation of the associated coding sequence. Regulatory sequences may include, and are not limited to, promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

“Promoter” refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an “enhancer” is a DNA sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoter sequences can also be located within the transcribed portions of genes, and/or downstream of the transcribed sequences. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of an isolated nucleic acid fragment in different tissues or cell types, or at

different stages of development, or in response to different environmental conditions. Promoters which cause an isolated nucleic acid fragment to be expressed in most cell types at most times are commonly referred to as “constitutive promoters”. New promoters of various types useful in plant cells are constantly being discovered; numerous examples may be found in the compilation by Okamuro and Goldberg, (1989) Biochemistry of Plants 15:1-82. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of some variation may have identical promoter activity.

Specific examples of promoters that may be useful in expressing the nucleic acid fragments of the invention include, but are not limited to, the oleosin promoter (PCT Publication WO99/65479, published December 12, 1999), the maize 27kD zein promoter (Ueda et al (1994) *Mol. Cell. Biol. 14*:4350-4359), the ubiquitin promoter (Christensen et al (1992) *Plant Mol. Biol. 18*:675-680), the SAM synthetase promoter (PCT Publication WO00/37662, published June 29, 2000), the CaMV 35S (Odell et al (1985) *Nature 313*:810-812), and the promoter described in PCT Publication WO02/099063 published December 12, 2002.

The “translation leader sequence” refers to a polynucleotide fragment located between the promoter of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner, R. and Foster, G. D. (1995) *Mol. Biotechnol. 3*:225-236).

An “intron” is an intervening sequence in a gene that does not encode a portion of the protein sequence. Thus, such sequences are transcribed into RNA but are then excised and are not translated. The term is also used for the excised RNA sequences.

The “3’ non-coding sequences” refer to DNA sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3’ end of the mRNA precursor. The

use of different 3' non-coding sequences is exemplified by Ingelbrecht, I. L., et al. (1989) *Plant Cell 1*:671-680.

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook, J., Fritsch, E.F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989. Transformation methods are well known to those skilled in the art and are described below.

"PCR" or "Polymerase Chain Reaction" is a technique for the synthesis of large quantities of specific DNA segments, consists of a series of repetitive cycles (Perkin Elmer Cetus Instruments, Norwalk, CT). Typically, the double stranded DNA is heat denatured, the two primers complementary to the 3' boundaries of the target segment are annealed at low temperature and then extended at an intermediate temperature. One set of these three consecutive steps is referred to as a cycle.

"Stable transformation" refers to the transfer of a nucleic acid fragment into a genome of a host organism, including nuclear and organellar genomes, resulting in genetically stable inheritance.

In contrast, "transient transformation" refers to the transfer of a nucleic acid fragment into the nucleus, or DNA-containing organelle, of a host organism resulting in gene expression without integration or stable inheritance.

Host organisms comprising the transformed nucleic acid fragments are referred to as "transgenic" organisms.

The term "amplified" means the construction of multiple copies of a nucleic acid sequence or multiple copies complementary to the nucleic acid sequence using at least one of the nucleic acid sequences as a template. Amplification systems include the polymerase chain reaction (PCR) system, ligase chain reaction (LCR) system, nucleic acid sequence based amplification (NASBA, Cangene, Mississauga, Ontario), Q-Beta Replicase systems, transcription-based amplification system (TAS), and strand displacement amplification (SDA). See, e.g., Diagnostic Molecular Microbiology: Principles and Applications, D. H. Persing et al., Ed., American Society for Microbiology, Washington, D.C. (1993). The product of amplification is termed an amplicon.

The term "chromosomal location" includes reference to a length of a chromosome which may be measured by reference to the linear segment of DNA which it comprises. The chromosomal location can be defined by reference to two unique DNA sequences, i.e., markers.

The term "marker" includes reference to a locus on a chromosome that serves to identify a unique position on the chromosome. A "polymorphic marker" includes reference to a marker which appears in multiple forms (alleles) such that different forms of the marker, when they are present in a homologous pair, allow transmission of each of the chromosomes in that pair to be followed. A genotype may be defined by use of one or a plurality of markers.

The present invention includes, *inter alia,* compositions and methods for altering or modulating (i.e., increasing or decreasing) the level of cytosolic pyrophosphatase polypeptides described herein in plants. The size of the oil, protein, starch and soluble carbohydrate pools in soybean seeds can be modulated or altered (i.e. increased or decreased) by altering the expression of a specific gene, encoding cytosolic pyrophosphatase (PPiase).

In one embodiment, the present invention concerns a transgenic plant comprising a recombinant DNA construct comprising a polynucleotide operably linked to at least one regulatory element, wherein said polynucleotide encodes a polypeptide having an amino acid sequence of at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 or 112 and wherein seed obtained from said transgenic plant has an altered oil, protein, starch and/or soluble carbohydrate content when compared to seed obtained from a control plant not comprising said recombinant DNA construct.

In a second embodiment the present invention concerns a transgenic seed obtained from the transgenic plant comprising a recombinant DNA construct comprising a polynucleotide operably linked to at least one regulatory element, wherein said polynucleotide encodes a polypeptide having an amino acid sequence of at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%,

82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 or 112 and wherein said transgenic seed has an altered oil, protein, starch and/or soluble carbohydrate content when compared to a control plant not comprising said recombinant DNA construct.

In a third embodiment the present invention concerns a transgenic seed obtained from the transgenic plant comprising a recombinant DNA construct comprising a polynucleotide operably linked to at least one regulatory element, wherein said polynucleotide encodes a polypeptide having an amino acid sequence of at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 or 112 and wherein said transgenic seed has an increased starch content of at least 0.5%, 1%, 1.5%, 2%, 2.5%, 3.0%, 3.5%, 4.0%, 4.5%, 5.0%, 5.5%, 6.0%, 6.5%, 7.0%, 7.5%, 8.0%, 8.5%, 9.0%, 9.5%, 10.0%, 10.5%, 11%, 11.5%, 12.0% 12.5%, 13.0, 13.5%. 14.0%, 14.5%, 15.0%, 15.5%, 15.0%, 16.5%, 17.0%, 17.5% 18.0%, 18.5%, 19.0%, 19.5%, 20.0%, 20.5%, 21.0%, 21.5%, 22.0%, 22.5%, 23.0%, 23.5%, 24.0%, 24.5%, 25.0%, 25.5%, 26.0%, 26.5%, 27.0%, 27.5%, 28.0%, 28.5%, 29%, 29.5%, 30.0%, 30.5%, 31.0%, 31.5%, 32.0%, 32.5%, 33.0%, 33.5%, 34.0%, 35.0%, 35.5%, 36.0%, 36.5%, 37.0%, 37.5%, 38.0%, 38.5%, 39.0%, 39.5%, 40.0%, 40.5%, 41.0%, 41.5%, 42.0%, 42.5%, 43.0%, 43.5%, 44.0%, 44.5%, 45.0%, 45.5%, 46.0%, 46.5%, 47.0%, 47.5%, 48.0%, 48.5%, 49.0%, 49.5%, or 50.0% on a dry weight basis when compared to a control seed not comprising said recombinant DNA construct.

In another embodiment, the present invention relates to a recombinant DNA construct comprising any of the isolated polynucleotides of the present invention operably linked to at least one regulatory sequence.

In another embodiment of the present invention, a recombinant construct of the present invention further comprises an enhancer.

In another embodiment, the present invention relates to a vector comprising any of the polynucleotides of the present invention.

In another embodiment, the present invention relates to an isolated polynucleotide fragment comprising a nucleotide sequence comprised by any of the polynucleotides of the present invention, wherein the nucleotide sequence contains at least 30, 40, 60, 100, 200, 300, 400, 500 or 600 nucleotides.

In another embodiment, the present invention relates to a method for transforming a cell comprising transforming a cell with any of the isolated polynucleotides of the present invention, and the cell transformed by this method. Advantageously, the cell is eukaryotic, e.g., a yeast or plant cell, or prokaryotic, e.g., a bacterium.

In yet another embodiment, the present invention relates to a method for transforming a cell, comprising transforming a cell with a polynucleotide of the present invention.

In another embodiment, the present invention relates to a method for producing a transgenic plant comprising transforming a plant cell with any of the isolated polynucleotides of the present invention and regenerating a transgenic plant from the transformed plant cell.

In another embodiment, a cell, plant, or seed comprising a recombinant DNA construct of the present invention.

In another embodiment, an isolated polynucleotide comprising: (i) a nucleic acid sequence encoding a polypeptide having an amino acid sequence of at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 or 112; or (ii) a full complement of the nucleic acid sequence of (i), wherein the full complement and the nucleic acid sequence of (i) consist of the same number of nucleotides and are 100% complementary. Any of the foregoing isolated polynucleotides may be utilized in any recombinant DNA constructs (including suppression DNA constructs) of the present invention. The polypeptide can be a PPiase or PPiase-like protein.

In another embodiment, an isolated polynucleotide comprising: (i) a nucleic acid sequence of at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95 or 111; or (ii) a full complement of the nucleic acid sequence of (i). Any of the foregoing isolated polynucleotides may be utilized in any recombinant DNA constructs (including suppression DNA constructs) of the present invention. The polypeptide can be a PPiase or PPiase-like protein.

In one aspect, the present invention includes recombinant DNA constructs (including suppression DNA constructs).

In another embodiment, the present invention relates to a method of selecting an isolated polynucleotide that alters, i.e. increases or decreases, the level of expression of a PPiase gene, protein or enzyme activity in a host cell, preferably a plant cell, the method comprising the steps of: (a) constructing an isolated polynucloetide of the present invention or an isolated recombinant DNA construct of the present invention; (b) introducing the isolated polynucleotide or the isolated recombinant DNA construct into a host cell; (c) measuring the level of the PPiase RNA, protein or enzyme activity in the host cell containing the isolated polynucloetide or recombinant DNA construct; (d) comparing the level of the PPiase RNA, protein or enzyme activity in the host cell containing the isolated poynucleotide or recombinant DNA construct with the level of the PPiase RNA, protein or enzyme activity in a host cell that does not contain the isolated polynucleotide or recombinant DNA construct, and selecting the isolated polynucleotide or recombinant DNA construct that alters , i.e., increases or decreases, the level of expression of the PPiase gene, protein or enzyme activity in the plant cell.

In another embodiment, this invention concerns a method for suppressing the level of expression of a gene encoding a cytosolic PPiase having PPiase activity in a transgenic plant, wherein the method comprises:

(a) transforming a plant cell with a fragment of the isolated polynucleotide of the invention;

(b) regenerating a transgenic plant from the transformed plant cell of 9a); and

(c) selecting a transgenic plant wherein the level of expression of a gene encoding a cytosolic polypeptide having PPiase activity has been suppressed.

Preferably, the gene encodes a cytosolic polypeptide having PPiase activity, and the plant is a soybean plant.

In another embodiment, the invention concerns a method for producing transgenic seed, the method comprising: a) transforming a plant cell with the recombinant DNA construct of (i) all or part of the nucleotide sequence set forth in SEQ ID NO: 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95 or 111, or (ii) the complement of (i); wherein (i) or (ii) is useful in co-suppression or antisense suppression of endogenous PPiase activity in a transgenic plant; (b) regenerating a transgenic plant from the transformed plant cell of (a); and (c) selecting a transgenic plant that produces transgenic seeds having an increase in oil content of at least 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, or 30 % compared to seed obtained from a non-transgenic plant. Preferably, the seed is a soybean plant.

In another embodiment, a plant comprising in its genome a recombinant DNA construct comprising: (a) a polynucleotide operably linked to at least one regulatory element, wherein said polynucleotide encodes a polypeptide having an amino acid sequence of at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 or 112 or (b) a suppression DNA construct comprising at least one regulatory element operably linked to: (i) all or part of: (A) a nucleic acid sequence encoding a polypeptide having an amino acid sequence of at least 70% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO:30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 or 112, or (B) a full complement of

the nucleic acid sequence of (b)(i)(A); or (ii) a region derived from all or part of a sense strand or antisense strand of a target gene of interest, said region having a nucleic acid sequence of at least 70% sequence identity, based on the Clustal V method of alignment, when compared to said all or part of a sense strand or antisense strand from which said region is derived, and wherein said target gene of interest encodes a cytosolic Pyrophosphatase, and wherein said plant has an altered oil, protein, starch and/or soluble carbohydrate content, when compared to a control plant not comprising said recombinant DNA construct.

A transgenic seed having an increased oil content of at least 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, or 30 % when compared to the oil content of a non-transgenic seed, wherein said transgenic seed comprises a recombinant DNA construct comprising: (a) all or part of the nucleotide sequence set forth in SEQ ID NO: 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95 or 111; or (b) the full-length complement of (a):wherein (a) or (b) is of sufficient length to inhibit expression of endogenous cytosolic pyrophosphatase activity in a transgenic plant and further wherein said seed has an increase in oil content of at least 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, or 30 % on a dry-weight basis, as compared to seed obtained from a non-transgenic plant.

Yet another embodiment of the invention concerns a transgenic seed comprising a recombinant DNA construct comprising:

(a) all or part of the nucleotide sequence set forth in SEQ ID NO: 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95 or 111; or (b) the full-length complement of (a): wherein (a) or (b) is of sufficient length to inhibit expression of endogenous cytosolic pyrophosphatase activity in a transgenic plant and further wherein said seed has an increase in oil content of at least 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, or 30 % on a dry-weight basis, as compared to seed obtained from a non-transgenic plant.

In another embodiment, the invention concerns a method for producing a transgenic plant, the method comprising: (a) transforming a plant cell with a recombinant DNA construct comprising a polynucleotide operably linked to at least one regulatory sequence, wherein the polynucleotide encodes a polypeptide having an amino acid sequence of at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 or 112; and (b) regenerating a plant from the transformed plant cell.

Another embodiment of the invention concerns, a method for producing transgenic seeds, the method comprising: (a) transforming a plant cell with a recombinant DNA construct comprising a polynucleotide operably linked to at least one regulatory sequence, wherein the polynucleotide encodes a polypeptide having an amino acid sequence of at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 or 112; and (b) regenerating a transgenic plant from the transformed plant cell of (a); and (c) selecting a transgenic plant that produces a transgenic seed having an altered oil, protein, starch and/or soluble carbohydrate content, as compared to a transgenic seed obtained from a non-transgenic plant.

Another embodiment of the invention concerns, a method for producing transgenic seeds, the method comprising: (a) transforming a plant cell with a recombinant DNA construct comprising a polynucleotide operably linked to at least one regulatory sequence, wherein the polynucleotide encodes a polypeptide having an amino acid sequence of at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80,

82, 84, 86, 88, 90, 92, 94, 96 or 112; and (b) regenerating a transgenic plant from the transformed plant cell of (a); and (c) selecting a transgenic plant that produces a transgenic seed having an increased starch content of at least 0.5%, 1%, 1.5%, 2%, 2.5%, 3.0%, 3.5%, 4.0%, 4.5%, 5.0%, 5.5%, 6.0%, 6.5%, 7.0%, 7.5%, 8.0%, 8.5%, 9.0%, 9.5%, 10.0%, 10.5%, 11%, 11.5%, 12.0% 12.5%, 13.0, 13.5%. 14.0%, 14.5%, 15.0%, 15.5%, 15.0%, 16.5%, 17.0%, 17.5% 18.0%, 18.5%, 19.0%, 19.5%, 20.0%, 20.5%, 21.0%, 21.5%, 22.0%, 22.5%, 23.0%, 23.5%, 24.0%, 24.5%, 25.0%, 25.5%, 26.0%, 26.5%, 27.0%, 27.5%, 28.0%, 28.5%, 29%, 29.5%, 30.0%, 30.5%, 31.0%, 31.5%, 32.0%, 32.5%, 33.0%, 33.5%, 34.0%, 35.0%, 35.5%, 36.0%, 36.5%, 37.0%, 37.5%, 38.0%, 38.5%, 39.0%, 39.5%, 40.0%, 40.5%, 41.0%, 41.5%, 42.0%, 42.5%, 43.0%, 43.5%, 44.0%, 44.5%, 45.0%, 45.5%, 46.0%, 46.5%, 47.0%, 47.5%, 48.0%, 48.5%, 49.0%, 49.5%, or 50.0% on a dry weight basis as compared to a transgenic seed obtained from a non-transgenic plant.

In another embodiment, the invention concerns a method for producing transgenic seed, the method comprising: (a) transforming a plant cell with a recombinant DNA construct comprising:(i) all or part of the nucleotide sequence set forth in SEQ ID NO: 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95 or 111; or (ii) the full-length complement of (i); wherein (i) or (ii) is of sufficient length to inhibit expression of endogenous cytosolic pyrophosphatase activity in a transgenic plant; (b) regenerating a transgenic plant from the transformed plant cell of (a); and (c) selecting a transgenic plant that produces a transgenic seed having an altered oil, protein, starch and/or soluble carbohydrate content, as compared to a transgenic seed obtained from a non-transgenic plant.

A method for producing transgenic seed, the method comprising: (a) transforming a plant cell with a recombinant DNA construct comprising: (i) all or part of the nucleotide sequence set forth in SEQ ID NO: 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95 or 111; or (ii) the full-length complement of (i);

wherein (i) or (ii) is of sufficient length to inhibit expression of endogenous cytosolic pyrophosphatase activity in a transgenic plant; (b) regenerating a transgenic plant from the transformed plant cell of (a); and (c) selecting a transgenic plant that produces a transgenic seed having an increase in oil content of at least 1%, 2%,

3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, or 30 %, on a dry-weight basis, as compared to a transgenic seed obtained from a non-transgenic plant.

Soybeans can be processed into a number of products. For example, “soy protein products” can include, and are not limited to, those items listed in Table 2. "Soy protein products".

TABLE 2

Soy Protein Products Derived from Soybean Seeds[a]

Whole Soybean Products	Processed Soy Protein Products
Roasted Soybeans	Full Fat and Defatted Flours
Baked Soybeans	Soy Grits
Soy Sprouts	Soy Hypocotyls
Soy Milk	Soybean Meal
	Soy Milk
	Soy Protein Isolates
Specialty Soy Foods/Ingredients	
Soy Milk	Soy Protein Concentrates
Tofu	Textured Soy Proteins
Tempeh	Textured Flours and Concentrates
Miso	Textured Concentrates
Soy Sauce	Textured Isolates
Hydrolyzed Vegetable Protein	
Whipping Protein	

[a]See Soy Protein Products: Characteristics, Nutritional Aspects and Utilization (1987). Soy Protein Council.

"Processing" refers to any physical and chemical methods used to obtain the products listed in Table A and includes, and is not limited to, heat conditioning, flaking and grinding, extrusion, solvent extraction, or aqueous soaking and extraction of whole or partial seeds. Furthermore, "processing" includes the methods used to concentrate and isolate soy protein from whole or partial seeds, as well as the various traditional Oriental methods in preparing fermented soy food products. Trading Standards and Specifications have been established for many of

these products (see National Oilseed Processors Association Yearbook and Trading Rules 1991-1992).

"White" flakes refer to flaked, dehulled cotyledons that have been defatted and treated with controlled moist heat to have a PDI (AOCS: Ba10-65) of about 85 to 90. This term can also refer to a flour with a similar PDI that has been ground to pass through a No. 100 U.S. Standard Screen size.

"Grits" refer to defatted, dehulled cotyledons having a U.S. Standard screen size of between No. 10 and 80.

"Soy Protein Concentrates" refer to those products produced from dehulled, defatted soybeans by three basic processes: acid leaching (at about pH 4.5), extraction with alcohol (about 55-80%), and denaturing the protein with moist heat prior to extraction with water. Conditions typically used to prepare soy protein concentrates have been described by Pass ((1975) U.S. Patent No. 3,897,574; Campbell et al., (1985) in New Protein Foods, ed. by Altschul and Wilcke, Academic Press, Vol. 5, Chapter 10, *Seed Storage Proteins*, pp 302-338).

"Extrusion" refers to processes whereby material (grits, flour or concentrate) is passed through a jacketed auger using high pressures and temperatures as a means of altering the texture of the material. "Texturing" and "structuring" refer to extrusion processes used to modify the physical characteristics of the material. The characteristics of these processes, including thermoplastic extrusion, have been described previously (Atkinson (1970) U.S. Patent No. 3,488,770, Horan (1985) In *New Protein Foods*, ed. by Altschul and Wilcke, Academic Press, Vol. 1A, Chapter 8, pp 367-414). Moreover, conditions used during extrusion processing of complex foodstuff mixtures that include soy protein products have been described previously (Rokey (1983) *Feed Manufacturing Technology III*, 222-237; McCulloch, U.S. Patent No. 4,454,804).

TABLE 3

Generalized Steps for Soybean Oil and Byproduct Production

Process Step	Process	Impurities Removed and/or By-Products Obtained
# 1	soybean seed	
# 2	oil extraction	meal

# 3	Degumming	lecithin
# 4	alkali or physical refining	gums, free fatty acids, pigments
# 5	water washing	soap
# 6	Bleaching	color, soap, metal
# 7	(hydrogenation)	
# 8	(winterization)	stearine
# 9	Deodorization	free fatty acids, tocopherols, sterols, volatiles
# 10	oil products	

More specifically, soybean seeds are cleaned, tempered, dehulled, and flaked, thereby increasing the efficiency of oil extraction. Oil extraction is usually accomplished by solvent (e.g., hexane) extraction but can also be achieved by a combination of physical pressure and/or solvent extraction. The resulting oil is called crude oil. The crude oil may be degummed by hydrating phospholipids and other polar and neutral lipid complexes that facilitate their separation from the nonhydrating, triglyceride fraction (soybean oil). The resulting lecithin gums may be further processed to make commercially important lecithin products used in a variety of food and industrial products as emulsification and release (i.e., antisticking) agents. Degummed oil may be further refined for the removal of impurities (primarily free fatty acids, pigments and residual gums). Refining is accomplished by the addition of a caustic agent that reacts with free fatty acid to form soap and hydrates phosphatides and proteins in the crude oil. Water is used to wash out traces of soap formed during refining. The soapstock byproduct may be used directly in animal feeds or acidulated to recover the free fatty acids. Color is removed through adsorption with a bleaching earth that removes most of the chlorophyll and carotenoid compounds. The refined oil can be hydrogenated, thereby resulting in fats with various melting properties and textures. Winterization (fractionation) may be used to remove stearine from the hydrogenated oil through

crystallization under carefully controlled cooling conditions. Deodorization (principally via steam distillation under vacuum) is the last step and is designed to remove compounds which impart odor or flavor to the oil. Other valuable byproducts such as tocopherols and sterols may be removed during the deodorization process. Deodorized distillate containing these byproducts may be sold for production of natural vitamin E and other high-value pharmaceutical products. Refined, bleached, (hydrogenated, fractionated) and deodorized oils and fats may be packaged and sold directly or further processed into more specialized products. A more detailed reference to soybean seed processing, soybean oil production, and byproduct utilization can be found in Erickson, Practical Handbook of Soybean Processing and Utilization, The American Oil Chemists' Society and United Soybean Board (1995). Soybean oil is liquid at room temperature because it is relatively low in saturated fatty acids when compared with oils such as coconut, palm, palm kernel, and cocoa butter.

For example, plant and microbial oils containing polyunsaturated fatty acids (PUFAs) that have been refined and/or purified can be hydrogenated, thereby resulting in fats with various melting properties and textures. Many processed fats (including spreads, confectionary fats, hard butters, margarines, baking shortenings, etc.) require varying degrees of solidity at room temperature and can only be produced through alteration of the source oil's physical properties. This is most commonly achieved through catalytic hydrogenation.

Hydrogenation is a chemical reaction in which hydrogen is added to the unsaturated fatty acid double bonds with the aid of a catalyst such as nickel. For example, high oleic soybean oil contains unsaturated oleic, linoleic, and linolenic fatty acids, and each of these can be hydrogenated. Hydrogenation has two primary effects. First, the oxidative stability of the oil is increased as a result of the reduction of the unsaturated fatty acid content. Second, the physical properties of the oil are changed because the fatty acid modifications increase the melting point resulting in a semi-liquid or solid fat at room temperature.

There are many variables which affect the hydrogenation reaction, which in turn alter the composition of the final product. Operating conditions including pressure, temperature, catalyst type and concentration, agitation, and reactor design are among the more important parameters that can be controlled. Selective

hydrogenation conditions can be used to hydrogenate the more unsaturated fatty acids in preference to the less unsaturated ones. Very light or brush hydrogenation is often employed to increase stability of liquid oils. Further hydrogenation converts a liquid oil to a physically solid fat. The degree of hydrogenation depends on the desired performance and melting characteristics designed for the particular end product. Liquid shortenings (used in the manufacture of baking products, solid fats and shortenings used for commercial frying and roasting operations) and base stocks for margarine manufacture are among the myriad of possible oil and fat products achieved through hydrogenation. A more detailed description of hydrogenation and hydrogenated products can be found in Patterson, H. B. W., Hydrogenation of Fats and Oils: Theory and Practice. The American Oil Chemists' Society (1994).

Hydrogenated oils have become somewhat controversial due to the presence of *trans*-fatty acid isomers that result from the hydrogenation process. Ingestion of large amounts of *trans*-isomers has been linked with detrimental health effects including increased ratios of low density to high density lipoproteins in the blood plasma and increased risk of coronary heart disease.

In another embodiment, the invention concerns a transgenic seed produced by any of the above methods. Preferably, the seed is a soybean seed.

The present invention concerns a transgenic soybean seed having increased total fatty acid content of at least 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, or 30% when compared to the total fatty acid content of a non-transgenic, null segregant soybean seed. It is understood that any measurable increase in the total fatty acid content of a transgenic versus a non-transgenic, null segregant would be useful. Such increases in the total fatty acid content would include, but are not limited to, at least 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, or 30%.

Regulatory sequences may include, and are not limited to, promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

"Tissue-specific" promoters direct RNA production preferentially in particular types of cells or tissues. Promoters which cause a gene to be expressed in most

cell types at most times are commonly referred to as “constitutive promoters”. New promoters of various types useful in plant cells are constantly being discovered; numerous examples may be found in the compilation by Okamuro and Goldberg (*Biochemistry of Plants* 15:1-82 (1989)). It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of some variation may have identical promoter activity.

A number of promoters can be used to practice the present invention. The promoters can be selected based on the desired outcome. The nucleic acids can be combined with constitutive, tissue-specific (preferred), inducible, or other promoters for expression in the host organism. Suitable constitutive promoters for use in a plant host cell include, for example, the core promoter of the Rsyn7 promoter and other constitutive promoters disclosed in WO 99/43838 and U.S. Patent No. 6,072,050; the core CaMV 35S promoter (Odell et al., *Nature* 313:810-812 (1985)); rice actin (McElroy et al., *Plant Cell* 2:163-171 (1990)); ubiquitin (Christensen et al., *Plant Mol. Biol.* 12:619-632 (1989) and Christensen et al., *Plant Mol. Biol.* 18:675-689 (1992)); pEMU (Last et al., *Theor. Appl. Genet.* 81:581-588 (1991)); MAS (Velten et al., EMBO J. 3:2723-2730 (1984)); ALS promoter (U.S. Patent No. 5,659,026), and the like. Other constitutive promoters include, for example, those discussed in U.S. Patent Nos. 5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680; 5,268,463; 5,608,142; and 6,177,611.

In choosing a promoter to use in the methods of the invention, it may be desirable to use a tissue-specific or developmentally regulated promoter. A tissue-specific or developmentally regulated promoter is a DNA sequence which regulates the expression of a DNA sequence selectively in particular cells/tissues of a plant. Any identifiable promoter may be used in the methods of the present invention which causes the desired temporal and spatial expression.

Promoters which are seed or embryo specific and may be useful in the invention include patatin (potato tubers) (Rocha-Sosa, M., et al. (1989) EMBO J. 8:23-29), convicilin, vicilin, and legumin (pea cotyledons) (Rerie, W.G., et al. (1991) Mol. Gen. Genet. 259:149-157; Newbigin, E.J., et al. (1990) Planta 180:461-470; Higgins, T.J.V., et al. (1988) Plant. Mol. Biol. 11:683-695), zein (maize endosperm) (Schemthaner, J.P., et al. (1988) EMBO J. 7:1249-1255), phaseolin (bean cotyledon) (Segupta-Gopalan, C., et al. (1985) Proc. Natl. Acad. Sci. U.S.A.

82:3320-3324), phytohemagglutinin (bean cotyledon) (Voelker, T. et al. (1987) EMBO J. 6:3571-3577), B-conglycinin and glycinin (soybean cotyledon) (Chen, Z-L, et al. (1988) EMBO J. 7:297- 302), glutelin (rice endosperm), hordein (barley endosperm) (Marris, C., et al. (1988) Plant Mol. Biol. 10:359-366), glutenin and gliadin (wheat endosperm) (Colot, V., et al. (1987) EMBO J. 6:3559-3564), and sporamin (sweet potato tuberous root) (Hattori, T., et al. (1990) Plant Mol. Biol. 14:595-604). Promoters of seed-specific genes operably linked to heterologous coding regions in chimeric gene constructions maintain their temporal and spatial expression pattern in transgenic plants. Such examples include *Arabidopsis thaliana* 2S seed storage protein gene promoter to express enkephalin peptides in *Arabidopsis* and *Brassica napus* seeds (Vanderkerckhove et al., Bio/Technology 7:L929-932 (1989)), bean lectin and bean beta-phaseolin promoters to express luciferase (Riggs et al., Plant Sci. 63:47-57 (1989)), and wheat glutenin promoters to express chloramphenicol acetyl transferase (Colot et al., EMBO J 6:3559- 3564 (1987)).

A plethora of promoters is described in WO 00/18963, published on April 6, 2000, the disclosure of which is hereby incorporated by reference. Examples of seed-specific promoters include, and are not limited to, the promoter for soybean Kunitz trypsin inhibitor (Kti3, Jofuku and Goldberg, *Plant Cell* 1:1079-1093 (1989)) β-conglycinin (Chen et al., *Dev. Genet.* 10:112-122 (1989)), the napin promoter, and the phaseolin promoter.

In some embodiments, isolated nucleic acids which serve as promoter or enhancer elements can be introduced in the appropriate position (generally upstream) of a non-heterologous form of a polynucleotide of the present invention so as to up or down regulate expression of a polynucleotide of the present invention. For example, endogenous promoters can be altered *in vivo* by mutation, deletion, and/or substitution (see, Kmiec, U.S. Patent No. 5,565,350; Zarling et al., PCT/US93/03868), or isolated promoters can be introduced into a plant cell in the proper orientation and distance from a cognate gene of a polynucleotide of the present invention so as to control the expression of the gene. Gene expression can be modulated under conditions suitable for plant growth so as to alter the total concentration and/or alter the composition of the polypeptides of the present invention in plant cell. Thus, the present invention includes compositions, and

methods for making, heterologous promoters and/or enhancers operably linked to a native, endogenous (i.e., non-heterologous) form of a polynucleotide of the present invention.

An intron sequence can be added to the 5' untranslated region or the coding sequence of the partial coding sequence to increase the amount of the mature message that accumulates in the cytosol. Inclusion of a spliceable intron in the transcription unit in both plant and animal expression constructs has been shown to increase gene expression at both the mRNA and protein levels up to 1000-fold (Buchman and Berg, *Mol. Cell Biol.* 8:4395-4405 (1988); Callis et al., *Genes Dev.* 1:1183-1200 (1987)). Such intron enhancement of gene expression is typically greatest when placed near the 5' end of the transcription unit. Use of maize introns Adh1-S intron 1, 2, and 6, the Bronze-1 intron are known in the art. See generally, *The Maize Handbook*, Chapter 116, Freeling and Walbot, Eds., Springer, New York (1994). A vector comprising the sequences from a polynucleotide of the present invention will typically comprise a marker gene which confers a selectable phenotype on plant cells. Typical vectors useful for expression of genes in higher plants are well known in the art and include vectors derived from the tumor-inducing (Ti) plasmid of *Agrobacterium tumefaciens* described by Rogers et al., *Meth. in Enzymol.* 153:253-277 (1987).

If polypeptide expression is desired, it is generally desirable to include a polyadenylation region at the 3'-end of a polynucleotide coding region. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA. The 3' end sequence to be added can be derived from, for example, the nopaline synthase or octopine synthase genes, or alternatively from another plant gene, or less preferably from any other eukaryotic gene.

Preferred recombinant DNA constructs include the following combinations: a) a nucleic acid fragment corresponding to a promoter operably linked to at least one nucleic acid fragment encoding a selectable marker, followed by a nucleic acid fragment corresponding to a terminator, b) a nucleic acid fragment corresponding to a promoter operably linked to a nucleic acid fragment capable of producing a stem-loop structure, and followed by a nucleic acid fragment corresponding to a terminator, and c) any combination of a) and b) above. Preferably, in the stem-loop structure at least one nucleic acid fragment that is capable of suppressing

expression of a native gene comprises the “loop” and is surrounded by nucleic acid fragments capable of producing a stem.

Preferred methods for transforming dicots and obtaining transgenic plants have been published, among others, for cotton (U.S. Patent No. 5,004,863, U.S. Patent No. 5,159,135); soybean (U.S. Patent No. 5,569,834, U.S. Patent No. 5,416,011); *Brassica* (U.S. Patent No. 5,463,174); peanut (Cheng et al. (1996) *Plant Cell Rep*. *15*:653-657, McKently et al. (1995) *Plant Cell Rep*. *14*:699-703); papaya (Ling, K. et al. (1991) Bio/technology 9:752-758); and pea (Grant et al. (1995) Plant *Cell Rep*. 15:254-258). For a review of other commonly used methods of plant transformation see Newell, C.A. (2000) *Mol. Biotechnol. 16*:53-65. One of these methods of transformation uses *Agrobacterium rhizogenes* (Tepfler, M. and Casse-Delbart, F. (1987) *Microbiol. Sci. 4*:24-28). Transformation of soybeans using direct delivery of DNA has been published using PEG fusion (PCT publication WO 92/17598), electroporation (Chowrira, G.M. et al. (1995) *Mol. Biotechnol. 3*:17-23; Christou, P. et al. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*:3962-3966), microinjection, or particle bombardment (McCabe, D.E. et. Al. (1988) *BiolTechnology* 6:923; Christou et al. (1988) *Plant Physiol*. 87:671-674).

There are a variety of methods for the regeneration of plants from plant tissue. The particular method of regeneration will depend on the starting plant tissue and the particular plant species to be regenerated. The regeneration, development and cultivation of plants from single plant protoplast transformants or from various transformed explants are well known in the art (Weissbach and Weissbach, (1988) In.: Methods for Plant Molecular Biology, (Eds.), Academic Press, Inc., San Diego, CA). This regeneration and growth process typically includes the steps of selection of transformed cells, culturing those individualized cells through the usual stages of embryonic development through the rooted plantlet stage. Transgenic embryos and seeds are similarly regenerated. The resulting transgenic rooted shoots are thereafter planted in an appropriate plant growth medium such as soil. The regenerated plants may be self-pollinated. Otherwise, pollen obtained from the regenerated plants is crossed to seed-grown plants of agronomically important lines. Conversely, pollen from plants of these important lines is used to pollinate regenerated plants. A transgenic plant of the present invention containing a desired polypeptide(s) is cultivated using methods well known to one skilled in the art.

In addition to the above discussed procedures, practitioners are familiar with the standard resource materials which describe specific conditions and procedures for the construction, manipulation and isolation of macromolecules (e.g., DNA molecules, plasmids, etc.), generation of recombinant DNA fragments and recombinant expression constructs and the screening and isolating of clones, (see for example, Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press; Maliga et al. (1995) Methods in Plant Molecular Biology, Cold Spring Harbor Press; Birren et al. (1998) Genome Analysis: Detecting Genes, 1, Cold Spring Harbor, New York; Birren et al. (1998) Genome Analysis: Analyzing DNA, 2, Cold Spring Harbor, New York; Plant Molecular Biology: A Laboratory Manual, eds. Clark, Springer, New York (1997)).

Assays to detect proteins may be performed by SDS-polyacrylamide gel electrophoresis or immunological assays. Assays to detect levels of substrates or products of enzymes may be performed using gas chromatography or liquid chromatography for separation and UV or visible spectrometry or mass spectrometry for detection, or the like. Determining the levels of mRNA of the enzyme of interest may be accomplished using northern-blotting or RT-PCR techniques. Once plants have been regenerated, and progeny plants homozygous for the transgene have been obtained, plants will have a stable phenotype that will be observed in similar seeds in later generations.

In another aspect, this invention includes a polynucleotide of this invention or a functionally equivalent subfragment thereof useful in antisense inhibition or cosuppression of expression of nucleic acid sequences encoding proteins having cytosolic pyrophosphatase activity, most preferably in antisense inhibition or cosuppression of an endogenous cytosolic pyrophosphatase gene.

Protocols for antisense inhibition or co-suppression are well known to those skilled in the art.

The sequences of the polynucleotide fragments used for suppression do not have to be 100% identical to the sequences of the polynucleotide fragment found in the gene to be suppressed. For example, suppression of all the subunits of the soybean seed storage protein ß-conglycinin has been accomplished using a polynucleotide derived from a portion of the gene encoding the α subunit (U.S. Patent No. 6,362,399). ß-conglycinin is a heterogeneous glycoprotein composed of

varying combinations of three highly negatively charged subunits identified as α, α' and β. The polynucleotide sequences encoding the α and α' subunits are 85% identical to each other while the polynucleotide sequences encoding the β subunit are 75 to 80% identical to the α and α' subunits, respectively. Thus, polynucleotides that are at least 75% identical to a region of the polynucleotide that is target for suppression have been shown to be effective in suppressing the desired target. The polynucleotide may be at least 80% identical, at least 90% identical, at least 95% identical, or about 100% identical to the desired target sequence.

The isolated nucleic acids and proteins and any embodiments of the present invention can be used over a broad range of plant types, particularly dicots such as the species of the genus *Glycine*.

It is believed that the nucleic acids and proteins and any embodiments of the present invention can be with monocots as well including, but not limited to, *Graminiae* including *Sorghum bicolor* and *Zea mays*.

The isolated nucleic acid and proteins of the present invention can also be used in species from the following dicot genera: *Cucurbita, Rosa, Vitis, Juglans, Fragaria, Lotus, Medicago, Onobrychis, Trifolium, Trigonella, Vigna, Citrus, Linum, Geranium, Manihot, Daucus, Arabidopsis, Brassica, Raphanus, Sinapis, Atropa, Capsicum, Datura, Hyoscyamus, Lycopersicon, Nicotiana, Solanum, Petunia, Digitalis, Majorana, Cichorium, Helianthus, Lactuca, Antirrhinum, Pelargonium, Ranunculus, Senecio, Salpiglossis, Cucumis, Browallia, Glycine, Pisum, Phaseolus,* and from the following monocot genera: *Bromus, Asparagus, Hemerocallis, Panicum, Pennisetum, Lolium, Oryza, Avena, Hordeum, Secale, Triticum, Bambusa*, *Dendrocalamus*, and *Melocanna*.

EXAMPLES

The present invention is further defined in the following Examples, in which parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions. Thus, various modifications of the invention in

addition to those shown and described herein will be apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims.

The disclosure of each reference set forth herein is incorporated herein by reference in its entirety.

EXAMPLE 1

Creation of an *Arabidopsis* Population with Activation-Tagged Genes

An 18.49-kb T-DNA based binary construct was created, pHSbarENDs2 (SEQ ID NO:1; FIG. 3), that contains four multimerized enhancer elements derived from the Cauliflower Mosaic Virus 35S promoter (corresponding to sequences -341 to -64, as defined by Odell et al., *Nature* 313:810-812 (1985)). The construct also contains vector sequences (pUC9) and a poly-linker (SEQ ID NO:2) to allow plasmid rescue, transposon sequences (Ds) to remobilize the T-DNA, and the bar gene to allow for glufosinate selection of transgenic plants. In principle, only the 10.8-kb segment from the right border (RB) to left border (LB) inclusive will be transferred into the host plant genome. Since the enhancer elements are located near the RB, they can induce cis-activation of genomic loci following T-DNA integration.

Arabidopsis activation-tagged populations were created by whole plant *Agrobacterium* transformation. The pHSbarENDs2 (SEQ ID NO:1) construct was transformed into *Agrobacterium tumefaciens* strain C58, grown in lysogeny broth medium at 25 °C to OD600 ~1.0. Cells were then pelleted by centrifugation and resuspended in an equal volume of 5% sucrose/0.05% Silwet L-77 (OSI Specialties, Inc). At early bolting, soil grown *Arabidopsis thaliana* ecotype Col-0 were top watered with the *Agrobacterium* suspension. A week later, the same plants were top watered again with the same *Agrobacterium* strain in sucrose/Silwet. The plants were then allowed to set seed as normal. The resulting T1 seed were sown on soil, and transgenic seedlings were selected by spraying with glufosinate (FINALE®; AgrEvo; Bayer Environmental Science). A total of 100,000 glufosinate resistant T1 seedlings were selected. T2 seed from each line was kept separate. Small aliquots of T2 seed from independently generated activation-tagged lines were pooled. The pooled seed were planted in soil and plants were grown to maturity producing T3 seed pools each comprised of seed derived from 96 activation-tagged lines.

EXAMPLE 2

Identification and characterization of mutant line lo15571

A method for screening Arabidopsis seed density was developed based on Focks and Benning (1998) with significant modifications. Arabidopsis seeds can be separated according to their density. Density layers were prepared by a mixture of 1,6 dibromohexane (d=1.6), 1-bromohexane (d=1.17) and mineral oil (d=0.84) at different ratios. From the bottom to the top of the tube, 6 layers of organic solvents each comprised of 2 mL were added sequentially. The ratios of 1,6 dibromohexane:1-bromohexane:mineral oil for each layer were 1:1:0, 1:2:0, 0:1:0, 0:5:1, 0:3:1, 0:0:1. About 600 mg of T3 seed of a given pool of 96 activation-tagged lines corresponding to about 30,000 seeds were loaded on to the surface layer of a 15 ml glass tube containing said step gradient. After centrifugation for 5 min at 2000 x g, seeds were separated according to their density. The seeds in the lower two layers of the step gradient and from the bottom of the tube were collected. Organic solvents were removed by sequential washing with 100 % and 80 % ethanol and seeds were sterilized using a solution of 5 % hypochloride (NaOCl) in water. Seed were rinsed in sterile water and plated on MS-1 media comprised of 0.5 x MS salts, 1% (W/V) sucrose, 0.05 MES/KOH (pH 5.8), 200 µg/mL, 10 g/L agar and 15 mg L^{-1} glufosinate ammonium (Basta; Sigma Aldrich, USA). A total of 520 T3 pools each derived from 96 T2 activation-tagged lines were screened in this manner. Seed pool 225 when subjected to density gradient centrifugation as described above produced about 20 seed with increased density. These seed were sterilized and plated on selective media containing Basta. Basta-resistant seedlings were transferred to soil and plants were grown in a controlled environment (22 °C, 16 h light/8 h dark, 100-200 µE $m^{-2}s^{-1}$). to maturity for about 8-10 weeks alongside three untransformed wild type plants of the Columbia ecotype. Oil content of T4 seed and control seed was measured by NMR as follows.

NMR based analysis of seed oil content:

Seed oil content was determined using a Maran Ultra NMR analyzer (Resonance Instruments Ltd, Whitney, Oxfordshire, UK). Samples (e.g., batches of Arabidopsis seed ranging in weight between 5 and 200 mg) were placed into pre-weighed 2 mL polypropylene tubes (Corning Inc, Corning NY, USA; Part no. 430917) previously labeled with unique bar code identifiers. Samples were then

placed into 96 place carriers and processed through the following series of steps by an ADEPT COBRA 600™ SCARA robotic system:

1. pick up tube (the robotic arm was fitted with a vacuum pickup devise);
2. read bar code;
3. expose tube to antistatic device (ensured that Arabidopsis seed were not adhering to the tube walls);
4. weigh tube (containing the sample), to 0.0001 g precision;
5. take NMR reading; measured as the intensity of the proton spin echo 1 msec after a 22.95 MHz signal had been applied to the sample (data was collected for 32 NMR scans per sample);
6. return tube to rack; and
7. repeat process with next tube.

Bar codes, tubes weights and NMR readings were recorded by a computer connected to the system. Sample weight was determined by subtracting the polypropylene tube weight from the weight of the tube containing the sample.

Seed oil content of soybeans seed was calculated as follows:

$$\% \text{ oil (\% wt basis)} = \frac{(\text{NMR signal} / \text{sample wt (g)}) - 70.58}{351.45}$$

Calibration parameters were determined by precisely weighing samples of soy oil (ranging from 0.0050 to 0.0700 g at approximately 0.0050 g intervals; weighed to a precision of 0.0001 g) into Corning tubes (see above) and subjecting them to NMR analysis. A calibration curve of oil content (% seed wt basis; assuming a standard seed weight of 0.1500 g) to NMR value was established.

The relationship between seed oil contents measured by NMR and absolute oil contents measured by classical analytical chemistry methods was determined as follows. Fifty soybean seed, chosen to have a range of oil contents, were dried at 40 °C in a forced air oven for 48 h. Individual seeds were subjected to NMR analysis, as described above, and were then ground to a fine powder in a GenoGrinder (SPEX Centriprep (Metuchen, N.J., U.S.A.); 1500 oscillations per minute, for 1 minute). Aliquots of between 70 and 100 mg were weighed (to 0.0001 g precision) into 13 x 100 mm glass tubes fitted with Teflon® lined screw caps; the

remainder of the powder from each bean was used to determine moisture content, by weight difference after 18 h in a forced air oven at 105 °C. Heptane (3 mL) was added to the powders in the tubes and after vortex mixing samples were extracted, on an end-over-end agitator, for 1h at room temperature. The extracts were centrifuged, 1500 x g for 10 min, the supernatant decanted into a clean tube and the pellets were extracted two more times (1 h each) with 1 mL heptane. The supernatants from the three extractions were combined and 50 µL internal standard (triheptadecanoic acid; 10 mg / mL toluene) was added prior to evaporation to dryness at room temperature under a stream of nitrogen gas; standards containing 0, 0.0050, 0.0100, 0.0150, 0.0200 and 0.0300 g soybean oil, in 5 mL heptane, were prepared in the same manner. Fats were converted to fatty acid methyl esters (FAMEs) by adding 1 mL 5% sulfuric acid (v:v. in anhydrous methanol) to the dried pellets and heating them at 80 °C for 30 min, with occasional vortex mixing. The samples were allowed to cool to room temperature and 1 mL 25% aqueous sodium chloride was added followed by 0.8 mL heptane. After vortex mixing the phases were allowed to separate and the upper organic phase was transferred to a sample vial and subjected to GC analysis.

Plotting NMR determined oil contents versus GC determined oil contents resulted in a linear relationship between 9.66 and 26.27% oil (GC values; % seed wt basis) with a slope of 1.0225 and an R^2 of 0.9744; based on a seed moisture content that averaged 2.6 +/- 0.8 %.

Seed oil content (on a % seed weight basis) of Arabidopsis seed was calculated as follows:

mg oil = (NMR signal - 2.1112)/37.514;

% oil = [(mg oil)/1000]/[g of seed sample weight] x 100.

Prior to establishing this formula, Arabidopsis seed oil was extracted as follows. Approximately 5 g of mature Arabidopsis seed (cv Columbia) were ground to a fine powder using a mortar and pestle. The powder was placed into a 33 x 94 mm paper thimble (Ahlstrom # 7100-3394; Ahlstrom, Mount Holly Springs, PA, USA) and the oil extracted during approximately 40 extraction cycles with petroleum ether (BP 39.9 – 51.7 °C) in a Soxhlet apparatus. The extract was allowed to cool and the crude oil was recovered by removing the solvent under vacuum in a rotary evaporator. Calibration parameters were determined by precisely weighing 11

standard samples of partially purified Arabidopsis oil (samples contained 3.6, 6.3, 7.9, 9.6 ,12.8, 16.3, 20.3, 28.2, 32.1, 39.9 and 60 mg of partially purified Arabidopsis oil) weighed to a precision of 0.0001 g) into 2 mL polypropylene tubes (Corning Inc, Corning NY, USA; Part no. 430917) and subjecting them to NMR analysis. A calibration curve of oil content (% seed weight basis) to NMR value was established.

Table 4 shows that the seed oil content of T4 activation-tagged line with Bar code ID K15571 is only 84 % of that of WT control plants grown in the same flat.

TABLE 4

Oil Content of T4 activation-tagged lines derived from T3 pool 225

BARCODE	% Oil	T3 pool ID #	oil content % of WT
K15557	42.8	225	97.0
K15558	43.0	225	97.3
K15559	44.3	225	100.4
K15560	42.7	225	96.8
K15561	43.8	225	99.2
K15562	42.5	225	96.3
K15563	43.2	225	98.0
K15564	43.0	225	97.4
K15565	43.3	225	98.1
K15566	43.6	225	98.9
K15567	42.1	225	95.4
K15568	39.2	225	88.7
K15569	43.3	225	98.0
K15570	43.2	225	97.8
K15571	**37.2**	**225**	**84.3**
K15572	41.9	225	95.0
K15573	42.9	225	97.2
K15574	43.2	225	98.0
K15575	42.9	225	97.2
K15582	43.3	225	98.2
K15583	43.5	WT	
K15585	44.6	WT	
K15586	44.3	WT	

K15571 was renamed lo15571. T4 seed were plated on selective media and a total of 10 glufosinate-resistant seedlings were planted in the same flat as four untransformed WT plants.

TABLE 5

Oil Content of T5 activation-tagged line lo15571

BARCODE	% Oil	T5 activation-tagged line ID	oil content % of WT
K22442	37.6	lo15571	86.9
K22448	37.4	lo15571	86.4
K22451	37.4	lo15571	86.4
K22447	37.1	lo15571	85.7
K22450	37.1	lo15571	85.7
K22445	36.9	lo15571	85.2
K22446	36.5	lo15571	84.3
K22443	36.1	lo15571	83.4
K22444	35.8	lo15571	82.7
K22449	30.0	lo15571	69.4
K22452	43.4	WT	
K22453	42.9	WT	
K22454	43.4	WT	
K22455	43.5	WT	

Table 5 shows that the seed oil content of T5 activation-tagged line lo15571 is between 69 and 87 % of that of WT control plants grown in the same flat. When plated on Basta-containing media all 10 T5 seed selections shown in Table 5 produced about 25 % of herbicide sensitive seedlings and 25 % of non-germinating seed. Applicants conclude that despite repeated selection on Basta containing media no lines homozygous for the lo15571-specific transgene could be recovered. It is believed that a gene that is important for development of viable seed was disrupted by the transgene insertion in lo15571. Twenty-four Basta-resistant T5 seedling of lo15571 were planted in the same flat alongside 12 untransformed WT control plants of the Columbia ecotype. Plants were grown to maturity and seed was bulk harvested from all 24 lo15571 and 12 WT plants. Oil content of lo15571 and WT seed was measured by NMR (Table 6). A total of four flats were grown and processed in this manner.

TABLE 6

Oil Content of T6 activation-tagged line lo15571

Barcode	% Oil	Seed ID	oil content % of WT
K35838	39.9	lo15571	89.9
K35839	44.4	WT	

K35761	36.8	lo15571	85.4
K35762	43.0	WT	
K35763	37.8	lo15571	88.2
K35764	42.8	WT	
K35765	37.0	lo15571	85.3
K35766	43.4	WT	

T6 seed of lo15571 and WT seed produced under identical conditions were subjected to compositional analysis as described below. Seed weight was measured by determining the weight of 100 seed. This analysis was performed in triplicate.

Tissue preparation:

Arabidopsis seed (approximately 0.5g in a ½ x 2" polycarbonate vial) was ground to a homogeneous paste in a GENOGRINDER® (3 x 30sec at 1400 strokes per minute, with a 15 sec interval between each round of agitation). After the second round of agitation the vials were removed and the Arabidopsis paste was scraped from the walls with a spatula prior to the last burst of agitation.

Determination of protein content:

Protein contents were estimated by combustion analysis on a Thermo FINNIGAN™ Flash 1112EA combustion analyzer running in the NCS mode (vanadium pentoxide was omitted) according to instructions of the manufacturer. Triplicate samples of the ground pastes, 4-8 mg, weighed to an accuracy of 0.001mg on a METTLER-TOLEDO® MX5 micro balance, were used for analysis. Protein contents were calculated by multiplying % N, determined by the analyzer, by 6.25. Final protein contents were expressed on a % tissue weight basis.

Determination of non-structural carbohydrate content:

Sub-samples of the ground paste were weighed (to an accuracy of 0.1mg) into 13x100mm glass tubes; the tubes had TEFLON® lined screw-cap closures. Three replicates were prepared for each sample tested.

Lipid extraction was performed by adding 2 ml aliquots of heptane to each tube. The tubes were vortex mixed and placed into an ultrasonic bath (VWR Scientific Model 750D) filled with water heated to 60 ºC. The samples were sonicated at full-power (~360 W) for 15 min and were then centrifuged (5 min x 1700 g). The supernatants were transferred to clean 13x100mm glass tubes and the

pellets were extracted 2 more times with heptane (2 ml, second extraction; 1 ml third extraction) with the supernatants from each extraction being pooled. After lipid extraction 1 ml acetone was added to the pellets and after vortex mixing, to fully disperse the material, they were taken to dryness in a Speedvac.

Non-structural carbohydrate extraction and analysis:

Two ml of 80% ethanol was added to the dried pellets from above. The samples were thoroughly vortex mixed until the plant material was fully dispersed in the solvent prior to sonication at 60 ºC for 15 min. After centrifugation, 5 min x 1700 g, the supernatants were decanted into clean 13x100mm glass tubes. Two more extractions with 80% ethanol were performed and the supernatants from each were pooled. The extracted pellets were suspended in acetone and dried (as above). An internal standard β-phenyl glucopyranoside (100 µl of a 0.5000 +/- 0.0010g/100ml stock) was added to each extract prior to drying in a Speedvac. The extracts were maintained in a desiccator until further analysis.

The acetone dried powders from above were suspended in 0.9 ml MOPS (3-N[Morpholino]propane-sulfonic acid; 50mM, 5mM $CaCl_2$, pH 7.0) buffer containing 100 U of heat-stable α-amylase (from *Bacillus licheniformis*; Sigma A-4551). Samples were placed in a heat block (90 ºC) for 75 min and were vortex mixed every 15 min. Samples were then allowed to cool to room temperature and 0.6 ml acetate buffer (285mM, pH 4.5) containing 5 U amyloglucosidase (Roche 110 202 367 001) was added to each. Samples were incubated for 15 –18 h at 55 ºC in a water bath fitted with a reciprocating shaker; standards of soluble potato starch (Sigma S-2630) were included to ensure that starch digestion went to completion.

Post-digestion the released carbohydrates were extracted prior to analysis. Absolute ethanol (6 ml) was added to each tube and after vortex mixing the samples were sonicated for 15 min at 60 ºC. Samples were centrifuged (5 min x 1700 g) and the supernatants were decanted into clean 13x100mm glass tubes. The pellets were extracted 2 more times with 3 ml of 80% ethanol and the resulting supernatants were pooled. Internal standard (100µl β-phenyl glucopyranoside, as above) was added to each sample prior to drying in a Speedvac.

Sample preparation and analysis:

The dried samples from the soluble and starch extractions described above were solubilized in anhydrous pyridine (Sigma-Aldrich P57506) containing 30 mg/ml

of hydroxylamine HCl (Sigma-Aldrich 159417). Samples were placed on an orbital shaker (300 rpm) overnight and were then heated for 1 hr (75 °C) with vigorous vortex mixing applied every 15 min. After cooling to room temperature, 1ml hexamethyldisilazane (Sigma-Aldrich H-4875) and 100 µl trifluoroacetic acid (Sigma-Aldrich T-6508) were added. The samples were vortex mixed and the precipitates were allowed to settle prior to transferring the supernatants to GC sample vials.

Samples were analyzed on an Agilent 6890 gas chromatograph fitted with a DB-17MS capillary column (15m x 0.32mm x 0.25um film). Inlet and detector temperatures were both 275 °C. After injection (2 µl, 20:1 split) the initial column temperature (150 °C) was increased to 180 °C at a rate of 3 °C/min and then at 25 °C/min to a final temperature of 320 °C. The final temperature was maintained for 10 min. The carrier gas was H_2 at a linear velocity of 51 cm/sec. Detection was by flame ionization. Data analysis was performed using Agilent ChemStation software. Each sugar was quantified relative to the internal standard and detector responses were applied for each individual carbohydrate (calculated from standards run with each set of samples). Final carbohydrate concentrations were expressed on a tissue weight basis.

Carbohydrates were identified by retention time matching with authentic samples of each sugar run in the same chromatographic set and by GC-MS with spectral matching to the NIST Mass Spectral Library Version 2a, build July 1 2002.

TABLE 7

Composition Analysis of lo15571 and WT Control Seed

Genotype	Bar code ID	Oil (%, NMR)	Protein %	Seed Weight (µg)	fructose (µg mg^{-1} seed)
lo15571	K35838	39.9	15.4	22.3	1.1
WT	K35839	44.4	14.7	19.0	0.2
	Δ TG/WT %	-10.1	+5.8	+17.4	+450

Genotype	Bar code ID	glucose (µg mg^{-1} seed)	sucrose (µg mg^{-1} seed)	raffinose (µg mg^{-1} seed)	stachyose (µg mg^{-1} seed)
lo15571	K35838	5.1	16.0	1.0	2.1
WT	K35839	3.5	12.5	0.8	1.6
	ΔTG/WT %	+45	+28	+25	+31

Genotype	Bar code ID	Oil (%,	Protein %	Seed	fructose

		NMR)		Weight (µg)	(µg mg^{-1} seed)
lo15571	K35761	36.8	17.4	21.0	3.7
WT	K35762	43.0	15.7	18.7	1.4
	Δ TG/WT %	-14.4	+11	+18.7	+164
Genotype	Bar code ID	glucose (µg mg^{-1} seed)	sucrose (µg mg^{-1} seed)	raffinose (µg mg^{-1} seed)	stachyose (µg mg^{-1} seed)
lo15571	K35761	11.1	18.4	1.3	2.8
WT	K35762	6.6	17.3	1.0	2.7
	ΔTG/WT %	+68	+18.4	+30	+3.7

Genotype	Bar code ID	Oil (%, NMR)	Protein %	Seed Weight (µg)	fructose (µg mg^{-1} seed)
lo15571	K35763	37.8	16.6	20.3	4.0
WT	K35764	42.8	16.2	21.0	1.1
	ΔTG/WT %	-11.7	+2.5	-3.3	+263
Genotype	Bar code ID	glucose (µg mg^{-1} seed)	sucrose (µg mg^{-1} seed)	raffinose (µg mg^{-1} seed)	stachyose (µg mg^{-1} seed)
lo15571	K35763	9.8	17.5	1.2	2.5
WT	K35764	6.8	16.9	1.2	3.2
	ΔTG/WT %	+44.1	+3.5	0	-19

Genotype	Bar code ID	Oil (%, NMR)	Protein %	Seed Weight (µg)	fructose (µg mg^{-1} seed)
lo15571	K35765	37.0	16.9	20.3	4.3
WT	K35766	43.4	15.8	16.7	1.1
	Δ TG/WT %	-14.7	+7	+21.5	+290
Genotype	Bar code ID	glucose (µg mg^{-1} seed)	sucrose (µg mg^{-1} seed)	raffinose (µg mg^{-1} seed)	stachyose (µg mg^{-1} seed)
lo15571	K35765	10.1	18.6	1.1	2.5
WT	K35766	6.3	17.2	0.8	2.5
	ΔTG/WT %	+60	+8.1	+37.5	0

The oil decrease in seed oil content of lo15571 is associated with an increase in seed weight and protein. The soluble carbohydrate profile of lo15571 differs from that of WT seed. The former shows a dramatic increase in fructose levels (1.6 – 4.5 fold increase compared to WT). There is also an increase in glucose levels and a small increase in sucrose levels associated with the presence of the lo15571 transgene (Table 7). A further characteristic of the lo15571 lines was a significant

increase in the sorbitol contents of the seed (data not shown). This indicates a perturbation of hexose metabolism in the tissues. The lo15571 line was crossed back to WT plants of the Columbia ecotype. To this end T6 seed of lo15571 were germinated on selective media containing glufosinate. Herbicide resistant seedlings were grown in soil. Pollen of lo15571 plant was used to fertilize emasculated immature flowers of WT plants. F1 seed were germinated on selective media, transferred to soil and 23 herbicide-resistant F1 plants were grown alongside four WT plants and four lo15571 plants in the same flat. WT seed were bulk harvested. F2 seed and lo15571 parent seed were harvested from individual plants. Table 8 shows that all 23 F_1 plants produced seed with an oil content that was lower than that of WT seed. The average decrease in seed oil content (compared to WT) of all F1 plant was 91.6 % which is very close to 90.2 % which was observed for the lo15571 parent.

TABLE 8

Seed oil content of F1 plants derived from a cross of lo15571 to WT plants of ecotype Columbia

Construct	BARCODE	% oil	oil content % of wt	avg. oil content % of WT
lo15571xCOL F_1	K40308	37.1	99.1	
	K40319	36.8	98.2	
	K40309	36.8	98.0	
	K40307	35.8	95.5	
	K40314	35.6	94.9	
	K40305	35.4	94.4	
	K40318	35.3	94.1	
	K40310	35.2	93.9	
	K40317	34.5	92.1	
	K40303	34.5	92.0	
	K40301	34.4	91.6	
	K40306	34.3	91.6	
	K40313	34.3	91.5	
	K40315	34.1	91.0	
	K40299	34.0	90.6	
	K40302	33.9	90.5	
	K40312	33.6	89.7	
	K40300	33.5	89.3	
	K40304	33.2	88.6	
	K40316	33.0	88.1	
	K40320	32.7	87.2	

	K40321	31.5	84.0	
	K40311	30.6	81.7	91.6
WT	K40322	37.5		
lo15571	K40327	34.5	92.1	
	K40325	34.4	91.7	
	K40324	33.9	90.4	
	K40326	33.7	90.0	
	K40323	32.6	87.0	90.2

In summary the lo15571 contains a single genetic locus that confers glufosinate herbicide resistance. Presence of this transgene is associated with a dominant low oil trait (reduction in oil content of 10-15 % compared to WT) that is accompanied by increased seed size, protein content and increased levels of fructose, glucose and sucrose in mature dry seed.

EXAMPLE 3

Identification of Activation-Tagged Genes

Genes flanking the T-DNA insert in the lo15571 lines were identified using one, or both, of the following two standard procedures: (1) thermal asymmetric interlaced (TAIL) PCR (Liu et al., *Plant J.* 8:457-63 (1995)); and (2) SAIFF PCR (Siebert et al., *Nucleic Acids Res.* 23:1087-1088 (1995)). In lines with complex multimerized T-DNA inserts, TAIL PCR and SAIFF PCR may both prove insufficient to identify candidate genes. In these cases, other procedures, including inverse PCR, plasmid rescue and/or genomic library construction, can be employed.

A successful result is one where a single TAIL or SAIFF PCR fragment contains a T-DNA border sequence and *Arabidopsis* genomic sequence. Once a tag of genomic sequence flanking a T-DNA insert is obtained, candidate genes are identified by alignment to publicly available *Arabidopsis* genome sequence. Specifically, the annotated gene nearest the 35S enhancer elements/T-DNA RB are candidates for genes that are activated.

To verify that an identified gene is truly near a T-DNA and to rule out the possibility that the TAIL/SAIFF fragment is a chimeric cloning artifact, a diagnostic PCR on genomic DNA is done with one oligo in the T-DNA and one oligo specific for the candidate gene. Genomic DNA samples that give a PCR product are interpreted as representing a T-DNA insertion. This analysis also verifies a situation in which more than one insertion event occurs in the same line, e.g., if multiple differing genomic fragments are identified in TAIL and/or SAIFF PCR analyses.

EXAMPLE 4

Identification of Activation-Tagged Genes in lo15571

Construction of pKR1478 for seed specific overexpression of genes in Arabidopsis

Plasmid pKR85 (SEQ ID NO:3; described in US Patent Application Publication US 2007/0118929 published on May 24,2007) was digested with HindIII and the fragment containing the hygromycin selectable marker was re-ligated together to produce pKR278 (SEQ ID NO:4).

Plasmid pKR407 (SEQ ID NO:5; described in PCT Int. Appl. WO 2008/124048 published on October 16, 2008) was digested with BamHI/HindIII and the fragment containing the Gy1 promoter/NotI/LegA2 terminator cassette was effectively cloned into the BamHI/HindIII fragment of pKR278 (SEQ ID NO:4) to produce pKR1468 (SEQ ID NO:6).

Plasmid pKR1468 (SEQ ID NO:6) was digested with NotI and the resulting DNA ends were filled using Klenow. After filling to form blunt ends, the DNA fragments were treated with calf intestinal alkaline phosphatase and separated using agarose gel electrophoresis. The purified fragment was ligated with cassette frmA containing a chloramphenicol resistance and ccdB genes flanked by attR1 and attR2 sites, using the Gateway® Vector Conversion System (Cat. No. 11823-029, Invitrogen Corporation) following the manufacturer's protocol to pKR1475 (SEQ ID NO:7).

Plasmid pKR1475 (SEQ ID NO:7) was digested with AscI and the fragment containing the Gy1 promoter/NotI/LegA2 terminator Gateway® L/R cloning cassette was cloned into the AscI fragment of binary vector pKR92 (SEQ ID NO:8; described in US Patent Application Publication US 2007/0118929 published on May 24,2007) to produce pKR1478 (SEQ ID NO:9).

In this way, genes flanked by attL1 and attL2 sites could be cloned into pKR1478 (SEQ ID NO:9) using Gateway® technology (Invitrogen Corporation) and the gene could be expressed in Arabidopsis from the strong, seed-specific soybean Gy1 promoter in soy.

The activation tagged-line (lo15571) showing reduced oil content was further analyzed. DNA from the line was extracted, and genes flanking the T-DNA insert in the mutant line were identified using ligation-mediated PCR (Siebert et al., *Nucleic Acids Res.* 23:1087-1088 (1995)). A single amplified fragment was identified that

contained a T-DNA border sequence and *Arabidopsis* genomic sequence. The sequence of this PCR product which contains part of the left border of the inserted T-DNA is set forth as SEQ ID NO:10. Once a tag of genomic sequence flanking a T-DNA insert was obtained, a candidate gene was identified by alignment to the completed *Arabidopsis* genome. Specifically, the SAIFF PCR product generated with PCR primers corresponding to the left border sequence of the T-DNA present in pHSbarENDs2 aligns with nucleotides 1221 – 1541 of the Arabidopsis gene At1g01040. The gene is also known as *DICER-like 1* (*DCL1*). Mutant alleles of this gene are known as *CARPEL FACTORY* (*CAF*), *SUSPENSOR 1* (*SUS1*), *SHORT INTEGUMENT 1* (*SIN1*), *ABNORMAL SUSPENSOR 1* (*ASU1*), *EMB76*, *EMB60*. The gene is annotated as an ATP-dependent helicase/ ribonuclease III with strong sequence similarity to the DICER class of proteins which act in miRNA processing. The DNA sequence generated using SAIFF and genomic DNA of lo15571 (SEQ ID NO:10) matches sequence of the first and second exon and first intron of At1g01040. Because of the location of the T-DNA in lo15571 we conclude that like the emb60 and emb70 alleles of *DCL1* the T-DNA insertion allele of *DCL1* present in lo15571 encodes a non-functional product of said gene which leads to embryo lethality. The low seed oil phenotype of herbicide resistant F1 plants that are heterozygous for the lo15571 transgene suggests that the disruption of At1g01040 is not related to the seed oil phenotype of lo15571.

<u>Validation of Candidate *Arabidopsis* Gene (At1g01050) via Transformation into *Arabidopsis*</u>

The gene At1g01050 is approximately 9 kb upstream of the SAIFF sequence corresponding to sequence adjacent to the left T-DNA border in lo15571. This gene is annotated as cytosolic, soluble pyrophosphatase it is also known as PPA1 and heterologous expression of PPA1 in *E. coli* confirms that this enzyme has pyrophosphatase activity (Navarro-De la Sancha, Ernesto; Coello-Coutino, Martha P.; Valencia-Turcotte, Lilian G.; Hernandez-Dominguez, Eric E.; Trejo-Yepes, Gisela; Rodriguez-Sotres, Rogelio. Characterization of two soluble inorganic pyrophosphatases from *Arabidopsis thaliana*. Plant Science (2007), 172(4), 796-807). Primers PPA1 FWD (SEQ ID NO:11) and PPA1 REV (SEQ ID NO:12) were used to amplify the At1g01050 ORF from applicants cDNA library of developing Arabidopsis seeds of the *erecta* mutant of the Landsberg ecotype. The PCR

product was cloned into pENTR (Invitrogen, USA) to give pENTR-PPA1 (SEQ ID NO:13). The PPA1 ORF was inserted in the sense orientation downstream of the GY1 promoter in binary plant transformation vector pKR1478 using Gateway LR recombinase (Invitrogen, USA) using manufacturer instructions. The sequence of the resulting plasmid pKR1478-PPA1 is set forth as SEQ ID NO:14.

pKR1478-PPA1 (SEQ ID NO:14) was introduced into *Agrobacterium tumefaciens* NTL4 (Luo et al, *Molecular Plant-Microbe Interactions* (2001) 14(1):98-103) by electroporation. Briefly, 1 µg plasmid DNA was mixed with 100 µL of electro-competent cells on ice. The cell suspension was transferred to a 100 µL electroporation cuvette (1 mm gap width) and electroporated using a BIORAD electroporator set to 1 kV, 400Ω and 25 µF. Cells were transferred to 1 mL LB medium and incubated for 2 h at 30 °C. Cells were plated onto LB medium containing 50 µg/mL kanamycin. Plates were incubated at 30 °C for 60 h. Recombinant Agrobacterium cultures (500 mL LB, 50 µg/mL kanamycin) were inoculated from single colonies of transformed agrobacterium cells and grown at 30 °C for 60 h. Cells were harvested by centrifugation (5000xg, 10 min) and resuspended in 1 L of 5 % (W/V) sucrose containing 0.05 % (V/V) Silwet. Arabidopsis plants were grown in soil at a density of 30 plants per 100 cm^2 pot in METRO-MIX® 360 soil mixture for 4 weeks (22 °C, 16 h light/8 h dark, 100 µE $m^{-2}s^{-1}$). Plants were repeatedly dipped into the Agrobacterium suspension harboring the binary vector pKR1478-PPA1 and kept in a dark, high humidity environment for 24 h. Post dipping, plants were grown for three to four weeks under standard plant growth conditions described above and plant material was harvested and dried for one week at ambient temperatures in paper bags. Seeds were harvested using a 0.425 mm mesh brass sieve.

Cleaned Arabidopsis seeds (2 grams, corresponding to about 100,000 seeds) were sterilized by washes in 45 mL of 80 % ethanol, 0.01 % TRITON® X-100, followed by 45 mL of 30 % (V/V) household bleach in water, 0.01 % TRITON® X-100 and finally by repeated rinsing in sterile water. Aliquots of 20,000 seeds were transferred to square plates (20 x 20 cm) containing 150 mL of sterile plant growth medium comprised of 0.5 x MS salts, 0.53 % (W/V) sorbitol, 0.05 MES/KOH (pH 5.8), 200 µg/mL TIMENTIN®, and 50 µg/mL kanamycin solidified with 10 g/L

agar. Homogeneous dispersion of the seed on the medium was facilitated by mixing the aqueous seed suspension with an equal volume of melted plant growth medium. Plates were incubated under standard growth conditions for ten days. Kanamycin-resistant seedlings were transferred to plant growth medium without selective agent and grown for one week before transfer to soil. Plants were grown to maturity and T2 seeds were harvested. Approximately 14 events were generated in this manner. A total of 42 Wild-type (WT) control plants were grown in the same flat and adjacent to flats containing pKR1478-PPA1 T1 plants. T2 seed were harvested and oil content was measured by NMR as described above.

TABLE 9

Seed oil content of T1 plants generated with binary vector pKR1478-PPA1 for seed specific over expression of At1g01050

Construct	BARCODE	% oil	oil content % of WT	avg. oil content % of WT
pKR1478-PPA1	K39596	38.2	104.0	
	K39597	37.7	102.7	
	K40552	34.4	93.8	
	K40549	33.9	92.4	
	K40550	33.9	92.3	
	K40551	33.8	92.1	
	K39595	33.6	91.4	
	K39594	33.5	91.2	
	K40548	33.1	90.3	
	K40547	33.0	90.0	
	K40545	32.5	88.6	
	K39593	32.1	87.4	
	K40544	31.6	86.1	
	K40546	27.3	74.4	91.2
WT		36.7		

Table 9 shows that the average seed oil content of all 14 T1 plants generated with pKR1478-PPA1 (SEQ ID NO:14) was 91.2 % of the oil content of Columbia control plants grown under identical conditions. Thus applicants have shown the seed specific over expression of At1g01050 leads to reduced seed oil content and moreover that the low oil phenotype of the lo15571 lines is most likely caused by increased expression of the At1g01050 gene resulting from the insertion of the 35S enhancer in the vicinity of the gene.

EXAMPLE 5

Seed-specific RNAi of At1g01050. Generation and phenotypic characterization of transgenic lines

A binary plant transformation vector pKR1482 (SEQ ID NO:15) for generation of hairpin constructs facilitating seed-specific RNAi was constructed. The RNAi related expression cassette that can be used for cloning of a given DNA fragment flanked by ATTL sites in sense and antisense orientation downstream of the GY1 promoter (see Example 4). The two gene fragments are interrupted by a sliceable intron sequence derived from the Arabidopsis gene At2g38080.

An intron of an Arabidopsis laccase gene (At2g38080) was amplified from genomic Arabidopsis DNA of ecotype Columbia using primers AthLcc IN FWD (SEQ ID NO:16) and AthLcc IN REV (SEQ ID NO:17). PCR products were cloned into pGEM T EASY (Promega, USA) according to manufacturer instructions and sequenced. The DNA sequence of the PCR product containing the laccase intron is set forth as SEQ ID NO:18. The PCR primers introduce an HpaI restriction site at the 5' end of the intron and restriction sites for NruI and SpeI at the 3' end of the intron. A three-way ligation of DNA fragments was performed as follows. XbaI digested, dephosphorylated DNA of pMBL18 (Nakano, Yoshio; Yoshida, Yasuo; Yamashita, Yoshihisa; Koga, Toshihiko. Construction of a series of pACYC-derived plasmid vectors. Gene (1995), 162(1), 157-8.) was ligated to the XbaI, EcoRV DNA fragment of PSM1318 (SEQ ID NO:19) containing ATTR12 sites a DNA Gyrase inhibitor gene (ccdB), a chloramphenicol acetyltransferase gene, an HpaI/SpeI restriction fragment excised from pGEM T EASY Lacc INT (SEQ ID NO:18) containing intron 1 of At2g38080. Ligation products were transformed into the DB 3.1 strain of *E. coli* (Invitrogen, USA). Recombinant clones were characterized by restriction digests and sequenced. The DNA sequence of the resulting plasmid pMBL18 ATTR12 INT is set forth as SEQ ID NO:20. DNA of pMBL18 ATTR12 INT was linearized with NruI, dephosphorylated and ligated to the XbaI, EcoRV DNA fragment of PSM1789 (SEQ ID NO: 21) containing ATTR12 sites and a DNA Gyrase inhibitor gene (ccdB). Prior to ligation ends of the PSM1789 restriction fragment had been filled in with T4 DNA polymerase (Promega, USA). Ligation products were transformed into the DB 3.1 strain of *E. coli* (Invitrogen, USA). Recombinant clones were characterized by restriction digests and sequenced. The

DNA sequence of the resulting plasmid pMBL18 ATTR12 INT ATTR21 is set forth as SEQ ID NO:22.

Plasmid pMBL18 ATTR12 INT ATTR21 (SEQ ID NO:22) was digested with XbaI and after filling to blunt the XbaI site generated, the resulting DNA was digested with Ecl136II and the fragment containing the attR cassettes was cloned into the NotI/BsiWI (where the NotI site was completely filled in) fragment of pKR1468 (SEQ ID NO:6), containing the Gy1 promoter, to produce pKR1480 (SEQ ID NO:23).

pKR1480 (SEQ ID NO:23) was digested with AscI and the fragment containing the Gy1 promoter/attR cassettes was cloned into the AscI fragment of binary vector pKR92 (SEQ ID NO:8) to produce pKR1482 (SEQ ID NO:15).

Primers PPA1 UTR FWD (SEQ ID NO:24) and PPA1 UTR REV (SEQ ID NO:25) were used to amplify the At1g01050 3'UTR from applicants cDNA library of developing Arabidopsis seeds of the *erecta* mutant of the Landsberg ecotype. The PCR product was cloned into pENTR (Invitrogen, USA) to give pENTR-PPA1 3'UTR (SEQ ID NO:26).

5 μg of plasmid DNA of pENTR-PPA1 3'UTR (SEQ ID NO:26) and pENTR-PPA1 (SEQ ID NO:13) was digested with EcoRV/HpaI. Restriction fragments of 528 bp (derived from pENTR-PPA1 3'UTR) and 955 bp (derived from pENTR-PPA1) were excised from agarose gels. Purified gene fragments containing ORF or 3'UTR sequences were inserted into vector pKR1482 using LR clonase (Invitrogen) according to the manufacturers instructions, to give pKR1482 PPA1 3'UTR (SEQ ID NO:27) or pKR1482 PPA1 ORF (SEQ ID NO:28).

pKR1482 PPA1 3'UTR (SEQ ID NO:27) or pKR1482 PPA1 ORF (SEQ ID NO:28) were introduced into *Agrobacterium tumefaciens* NTL4 (Luo et al, *Molecular Plant-Microbe Interactions* (2001) 14(1):98-103) by electroporation. Briefly, 1 μg plasmid DNA was mixed with 100 μL of electro-competent cells on ice. The cell suspension was transferred to a 100 μL electroporation cuvette (1 mm gap width) and electroporated using a BIORAD electroporator set to 1 kV, 400Ω and 25 μF. Cells were transferred to 1 mL LB medium and incubated for 2 h at 30 °C. Cells were plated onto LB medium containing 50 μg/mL kanamycin. Plates were incubated at 30 °C for 60 h. Recombinant Agrobacterium cultures (500 mL LB, 50 μg/mL kanamycin) were inoculated from single colonies of transformed

agrobacterium cells and grown at 30 °C for 60 h. Cells were harvested by centrifugation (5000xg, 10 min) and resuspended in 1 L of 5 % (W/V) sucrose containing 0.05 % (V/V) Silwet. Arabidopsis plants were grown in soil at a density of 30 plants per 100 cm^2 pot in METRO-MIX® 360 soil mixture for 4 weeks (22 °C, 16 h light/8 h dark, 100 µE $m^{-2}s^{-1}$). Plants were repeatedly dipped into the Agrobacterium suspension harboring the binary vectors pKR1482 PPA1 3'UTR (SEQ ID NO:27) or pKR1482 PPA1 ORF (SEQ ID NO:28) and kept in a dark, high humidity environment for 24 h. Plants were grown for three to four weeks under standard plant growth conditions described above and plant material was harvested and dried for one week at ambient temperatures in paper bags. Seeds were harvested using a 0.425 mm mesh brass sieve.

Cleaned Arabidopsis seeds (2 grams, corresponding to about 100,000 seeds) were sterilized by washes in 45 mL of 80 % ethanol, 0.01 % TRITON® X-100, followed by 45 mL of 30 % (V/V) household bleach in water, 0.01 % TRITON® X-100 and finally by repeated rinsing in sterile water. Aliquots of 20,000 seeds were transferred to square plates (20 x 20 cm) containing 150 mL of sterile plant growth medium comprised of 0.5 x MS salts, 0.53 % (W/V) sorbitol, 0.05 MES/KOH (pH 5.8), 200 µg/mL TIMENTIN®, and 50 µg/mL kanamycin solidified with 10 g/L agar. Homogeneous dispersion of the seed on the medium was facilitated by mixing the aqueous seed suspension with an equal volume of melted plant growth medium. Plates were incubated under standard growth conditions for ten days. Kanamycin-resistant seedlings were transferred to plant growth medium without selective agent and grown for one week before transfer to soil. Plants were grown to maturity and T2 seeds were harvested. A total of 25 and 60 events were generated with pKR1482 PPA1 ORF and pKR1482 PPA1 3'UTR, respectively. A total of 42 Wild-type (WT) control plants were grown in the same flat and adjacent to flats of pKR1482 PPA1 ORF and pKR1482 PPA1 3'UTR containing T1 plants. WT seeds and T2 seeds of transgenic lines were harvested and oil content was measured by NMR as described above.

TABLE 10

Seed oil content of T1 plants generated with binary vectors pKR1482-PPA1 and pKR1482-PPA1 3'UTR for seed specific gene suppression of At1g01050

Construct	BARCODE	% oil	oil content % of WT	avg. oil content % of WT
pKR1482 PPA1 ORF	C34251	43.7	119.1	
	C34257	43.0	117.1	
	C34246	42.9	117.0	
	C34242	42.5	115.9	
	C34241	41.3	112.5	
	C34248	40.6	110.6	
	C34252	40.4	110.2	
	C34256	40.3	109.9	
	C34264	40.1	109.3	
	C34258	40.1	109.2	
	C34255	39.9	108.7	
	C34260	39.9	108.6	
	C34253	39.2	106.9	
	C34262	38.8	105.9	
	C34263	38.8	105.8	
	C34240	38.7	105.3	
	C34244	38.3	104.4	
	C34250	38.3	104.3	
	C34254	38.0	103.4	
	C34249	37.9	103.3	
	C34245	36.9	100.4	
	C34261	36.3	98.8	
	C34247	34.6	94.4	
	C34259	33.7	91.9	
	C34243	26.5	72.1	105.8
WT		36.7		
pKR1482 PPA1 3'UTR	C34317	44.2	120.5	
	C34306	43.4	118.3	
	K40484	43.4	118.2	
	C34316	43.1	117.4	
	C34314	42.5	115.9	
	K40475	42.3	115.1	
	K40491	42.0	114.5	
	C34335	41.8	113.8	
	C34329	41.7	113.5	
	C34328	41.5	113.2	
	C34330	41.5	113.2	

	K40489	41.2	112.3	
	C34331	41.1	112.1	
	C34311	41.0	111.7	
	K40480	41.0	111.6	
	C34312	40.6	110.6	
	C34308	40.5	110.4	
	K40477	40.5	110.3	
	K40497	40.2	109.6	
	C34318	40.1	109.3	
	K40501	39.8	108.5	
	C34324	39.8	108.5	
	C34320	39.8	108.3	
	K40481	39.5	107.6	
	K40502	39.5	107.6	
	K40479	39.4	107.4	
	K40495	39.3	107.2	
	K40473	38.9	106.1	
	K40482	38.8	105.7	
	C34334	38.7	105.5	
	K40496	38.6	105.2	
	K40486	38.6	105.1	
	C34327	38.5	104.9	
	K40500	38.5	104.9	
	K40499	38.4	104.6	
	C34319	38.0	103.5	
	C34323	37.8	103.0	
	K40494	37.8	102.9	
	C34322	37.7	102.8	
	K40488	37.7	102.8	
	C34310	37.6	102.6	
	K40487	37.6	102.4	
	C34307	37.5	102.2	
	C34321	37.2	101.3	
	C34309	37.1	101.2	
	K40474	36.5	99.5	
	C34315	36.5	99.5	
	K40493	36.5	99.3	
	C34313	36.0	98.2	
	K40476	35.9	97.8	
	C34326	35.6	96.9	
	K40498	35.0	95.5	
	K40490	34.8	94.8	
	K40492	34.8	94.8	

	K40483	33.0	89.8	
	C34333	32.4	88.4	
	K40478	30.2	82.3	
	C34325	28.7	78.3	
	C34332	25.6	69.8	
	K40485	22.3	60.8	104.0
WT		36.7		

Table 10 shows that seed-specific down regulation of At1g01050 leads to increased oil content in Arabidopsis seed.

T2 seed of events C34251 and C34317 that carry transgenes pKR1482 PPA1 ORF and pKR1482 PPA1 3'UTR, respectively were plated on plant growth media containing kanamycin. For event C34251 and event C34317 24 and 12 kanamycin-resistant T2 seedlings, respectively, were grown to maturity alongside a total of 48 WT plants of the Columbia ecotype grown in the same or adjacent flats in the same growth chamber. Oil content of T3 seed is depicted in Table 11. Table 11 demonstrates that the oil increase associated with seed-specific down regulation of At1g01050 is heritable.

TABLE 11

Seed oil content of T2 plants generated with binary vectors pKR1482-PPA1 and pKR1482-PPA1 3'UTR for seed specific gene suppression of At1g01050

Construct	Event	T2 plant #	%oil	Oil content % of wt	Avg. oil content % of wt
pKR1482 PPA1 ORF	C3451	1	44.2	104.7	
		2	43.9	104.1	
		3	43.9	103.9	
		4	43.9	103.9	
		5	43.8	103.9	
		6	43.8	103.8	
		7	43.7	103.7	
		8	43.7	103.5	
		9	43.7	103.5	

		10	43.5	103.2	
		11	43.5	103.2	
		12	43.5	103.1	
		13	43.4	103.0	
		14	43.4	102.9	
		15	43.3	102.6	
		16	43.3	102.6	
		17	43.3	102.5	
		18	43.3	102.5	
		19	43.0	101.9	
		20	42.8	101.5	
		21	42.8	101.4	
		22	42.7	101.3	
		23	42.6	101.0	
		24	41.1	97.4	102.7
wt			42.2		
pKR1482 PPA1 3'UTR	C34317	1	43.5	103.2	
		2	43.5	103.0	
		3	43.4	102.8	
		4	43.3	102.7	
		5	43.3	102.6	
		6	43.2	102.4	
		7	43.1	102.1	
		8	43.1	102.1	
		9	43.0	101.9	
		10	42.5	100.7	
		11	42.5	100.7	
		12	42.2	100.0	102.0
wt			42.2		

EXAMPLE 6

Identification of genes of *Arabidopsis thaliana* closely related to At1g01050

Public DNA sequences (Arabidopsis Predicted Transcripts - TAIR8 (N) Gene sequences (predicted transcripts) from TAIR8 release, including mitochondrial and chloroplast-encoded genes (includes UTRs but not introns) were searched using the predicted amino acid sequence of At1g01050 and tBLASTn. There are four additional genes which share at least 71.9 % sequence identity to At1g01050. These genes and their properties and SEQ ID NOs are listed in Table 12.

TABLE 12

Arabidopsis genes closely related to At1g01050

Gene name	% AA sequence identity to At1g01050 (ClustalW)	SEQ ID NO: NT	SEQ ID NO: AA
PPA1/At1g01050	**100**	**29**	**30**
PPA2/At2g18230	**71.8**	**31**	**32**
PPA3/At2g46860	**88.7**	**33**	**34**
PPA4/At3g53620	**79.3**	**35**	**36**
PPA5/At4g01480	**91.1**	**37**	**38**

EXAMPLE 7

Identification of genes of *Brassica napus* closely-related to At1g01050

Public DNA sequences (NCBI and Brassica napus EST assembly (N) Brassica napus EST assembly version 3.0 (July 30, 2007) from the Gene Index Project at Dana-Farber Cancer Institute were searched using the predicted amino acid sequence of At1g01050 and tBLASTn. The assembly encompasses about 558465 public ESTs and has a total of 90310 sequences (47591 assemblies and 42719 singletons). There are a total of 11 genes which share at least 72.3 % amino acid sequence identity to At1g01050. These genes, their % identity to At1g01050 and SEQ ID NOs are listed in Table 13.

TABLE 13

Brassica napus genes closely related to At1g01050

Gene name	% AA sequence identity to At1g01050	SEQ ID NO: NT	SEQ ID NO: AA
TC23077	**88.3**	**39**	**40**
TC20341	**93.9**	**41**	**42**
TC16648	**93.4**	**43**	**44**
TC20135	**97.6**	**45**	**46**
TC23373	**89.2**	**47**	**48**
DY022345.1	**72.3**	**49**	**50**
TC34086	**82.2**	**51**	**52**
TC22517	**98.1**	**53**	**54**
TC56550	**72.3**	**55**	**56**
TC26534	**82.2**	**57**	**58**
TC16649	**97.6**	**59**	**60**

EXAMPLE 8

Identification of genes of soybean (*Glycine max*) closely-related to At1g01050

Public DNA sequences (Soybean cDNAs Glyma1.01 (JGI) (N) Predicted cDNAs from Soybean JGI Glyma1.01 genomic sequence, FGENESH predictions, and EST PASA analysis.) were searched using the predicted amino acid sequence of At1g01050 and tBLASTn. There are a total of 7 genes which share at least 77.5 % amino acid sequence identity At1g01050. These genes, their properties and SEQ ID NOs are listed in Table 14.

TABLE 14

Soybean genes closely related to At1g01050

Gene name	% AA sequence identity to At1g01050	SEQ ID NO: NT	SEQ ID NO: AA
Glyma19g35710	77.5	61	62
Glyma01g37790	78.4	63	64
Glyma03g33000	77.5	65	66
Glyma07g05390	89.7	67	68
Glyma10g05130	80.3	69	70
Glyma11g07530	78.9	71	72

EXAMPLE 9

Identification of genes of maize (*Zea mays*) closely-related to At1g01050

An assembly of proprietary and public maize EST DNA sequences (UniCorn 7.0 (N) Corn UniGene dataset, July 2007) was searched using the predicted amino acid sequence of At1g01050 and tBLASTn. There are a total of 5 genes which share at least 79.0% amino acid sequence identity to At1g01050. These genes, their properties and SEQ ID NOs are listed in Table 15.

TABLE 15

Maize genes closely related to At1g01050

Gene name	% AA sequence identity to At1g01050	SEQ ID NO: NT	SEQ ID NO: AA
PCO593895	80.8	73	74
PCO598466	84	75	76
PCO640614	84	77	78
PCO640979	84.9	79	80
PCO650999	79	81	82

EXAMPLE 10

Identification of genes of rice (*Oryza sativa*) closely-related to At1g01050

A public database of transcripts from rice gene models (Oryza sativa (japonica cultivar-group) MSU Rice Genome Annotation Project Osa1 release 6 (January 2009)) which includes untranslated regions (UTR) but no introns was searched

using the predicted amino acid sequence of At1g01050 and tBLASTn. There are a total of 7 genes which share at least 77.0% amino acid sequence identity to At1g01050. These genes, their properties and SEQ ID NOs are listed in Table 16.

TABLE 16

Rice genes closely related to At1g01050

Gene name	% AA sequence identity to At1g01050	SEQ ID NO: NT	SEQ ID NO: AA
LOC_Os10g26600.1	85.4	83	84
LOC_Os02g47600.1	77	85	86
LOC_Os05g02310.1	80.3	87	88
LOC_Os01g64670.1	80.7	89	90
LOC_Os04g59040.1	83.1	91	92
LOC_Os01g74350.1	78.4	93	94
LOC_Os05g36260.1	84	95	96

EXAMPLE 11

Expression of Chimeric Genes in Monocot Cells

A chimeric gene comprising a cDNA encoding the instant polypeptides in sense orientation with respect to the maize 27 kD zein promoter that is located 5' to the cDNA fragment, and the 10 kD zein 3' end that is located 3' to the cDNA fragment, can be constructed. The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites (NcoI or SmaI) can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the digested vector pML103 as described below. Amplification is then performed in a standard PCR. The amplified DNA is then digested with restriction enzymes NcoI and SmaI and fractionated on an agarose gel. The appropriate band can be isolated from the gel and combined with a 4.9 kb NcoI-SmaI fragment of the plasmid pML103. Plasmid pML103 has been deposited under the terms of the Budapest Treaty at ATCC (American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209), and bears accession number ATCC 97366. The DNA segment from pML103 contains a 1.05 kb SalI-NcoI promoter fragment of the maize 27 kD zein gene and a 0.96 kb SmaI-SalI fragment from the 3' end of the maize 10 kD zein gene in the vector pGem9Zf(+) (Promega). Vector and insert DNA can be ligated at 15°C overnight, essentially as described (Maniatis). The ligated DNA may then be used to transform *E. coli* XL1-Blue (Epicurian Coli XL-1 Blue™; Stratagene). Bacterial transformants can be screened by restriction enzyme

digestion of plasmid DNA and limited nucleotide sequence analysis using the dideoxy chain termination method (Sequenase ™ DNA Sequencing Kit; U.S. Biochemical). The resulting plasmid construct would comprise a chimeric gene encoding, in the 5' to 3' direction, the maize 27 kD zein promoter, a cDNA fragment encoding the instant polypeptides, and the 10 kD zein 3' region.

The chimeric gene described above can then be introduced into corn cells by the following procedure. Immature corn embryos can be dissected from developing caryopses derived from crosses of the inbred corn lines H99 and LH132. The embryos are isolated 10 to 11 days after pollination when they are 1.0 to 1.5 mm long. The embryos are then placed with the axis-side facing down and in contact with agarose-solidified N6 medium (Chu et al. (1975) *Sci. Sin. Peking* 18:659-668). The embryos are kept in the dark at 27°C. Friable embryogenic callus consisting of undifferentiated masses of cells with somatic proembryoids and embryoids borne on suspensor structures proliferate from the scutellum of these immature embryos. The embryogenic callus isolated from the primary explant can be cultured on N6 medium and sub-cultured on this medium every 2 to 3 weeks.

The plasmid, p35S/Ac (obtained from Dr. Peter Eckes, Hoechst Ag, Frankfurt, Germany) may be used in transformation experiments in order to provide for a selectable marker. This plasmid contains the *Pat* gene (see European Patent Publication 0 242 236) which encodes phosphinothricin acetyl transferase (PAT). The enzyme PAT confers resistance to herbicidal glutamine synthetase inhibitors such as phosphinothricin. The *pat* gene in p35S/Ac is under the control of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*.

The particle bombardment method (Klein et al. (1987) *Nature* 327:70-73) may be used to transfer genes to the callus culture cells. According to this method, gold particles (1 µm in diameter) are coated with DNA using the following technique. Ten µg of plasmid DNAs are added to 50 µL of a suspension of gold particles (60 mg per mL). Calcium chloride (50 µL of a 2.5 M solution) and spermidine free base (20 µL of a 1.0 M solution) are added to the particles. The suspension is vortexed during the addition of these solutions. After 10 minutes, the tubes are briefly centrifuged (5 sec at 15,000 rpm) and the supernatant removed. The

particles are resuspended in 200 µL of absolute ethanol, centrifuged again and the supernatant removed. The ethanol rinse is performed again and the particles resuspended in a final volume of 30 µL of ethanol. An aliquot (5 µL) of the DNA-coated gold particles can be placed in the center of a Kapton™flying disc (Bio-Rad Labs). The particles are then accelerated into the corn tissue with a Biolistic™ PDS-1000/He (Bio-Rad Instruments, Hercules CA), using a helium pressure of 1000 psi, a gap distance of 0.5 cm and a flying distance of 1.0 cm.

For bombardment, the embryogenic tissue is placed on filter paper over agarose-solidified N6 medium. The tissue is arranged as a thin lawn and covered a circular area of about 5 cm in diameter. The petri dish containing the tissue can be placed in the chamber of the PDS-1000/He approximately 8 cm from the stopping screen. The air in the chamber is then evacuated to a vacuum of 28 inches of Hg. The macrocarrier is accelerated with a helium shock wave using a rupture membrane that bursts when the He pressure in the shock tube reaches 1000 psi. Seven days after bombardment the tissue can be transferred to N6 medium that contains gluphosinate (2 mg per liter) and lacks casein or proline. The tissue continues to grow slowly on this medium. After an additional 2 weeks the tissue can be transferred to fresh N6 medium containing gluphosinate. After 6 weeks, areas of about 1 cm in diameter of actively growing callus can be identified on some of the plates containing the glufosinate-supplemented medium. These calli may continue to grow when sub-cultured on the selective medium.

Plants can be regenerated from the transgenic callus by first transferring clusters of tissue to N6 medium supplemented with 0.2 mg per liter of 2,4-D. After two weeks the tissue can be transferred to regeneration medium (Fromm et al. (1990) *Bio/Technology 8*:833-839).

EXAMPLE 12

Expression of Chimeric Genes in Dicot Cells

A seed-specific construct composed of the promoter and transcription terminator from the gene encoding the β subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris* (Doyle et al. (1986) *J. Biol. Chem. 261*:9228-9238) can be used for expression of the instant polypeptides in transformed soybean. The phaseolin construct includes about 500 nucleotides upstream (5') from the translation initiation codon and about 1650 nucleotides

downstream (3') from the translation stop codon of phaseolin. Between the 5' and 3' regions are the unique restriction endonuclease sites Nco I (which includes the ATG translation initiation codon), Sma I, Kpn I and Xba I. The entire construct is flanked by Hind III sites.

The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the expression vector. Amplification is then performed as described above, and the isolated fragment is inserted into a pUC18 vector carrying the seed construct.

Soybean embryos may then be transformed with the expression vector comprising sequences encoding the instant polypeptides. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from surface sterilized, immature seeds of the soybean cultivar A2872 can be cultured in the light or dark at 26°C on an appropriate agar medium for 6-10 weeks. Somatic embryos which produce secondary embryos are then excised and placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos which multiplied as early, globular staged embryos, the suspensions are maintained as described below. Soybean embryogenic suspension cultures can be maintained in 35 mL of liquid media on a rotary shaker, 150 rpm, at 26°C with fluorescent lights on a 16:8 hour day/night schedule. Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 mL of liquid medium.

Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Klein et al. (1987) *Nature* (London) *327*:70-73, U.S. Patent No. 4,945,050). A DuPont Biolistic™ PDS1000/HE instrument (helium retrofit) can be used for these transformations.

A selectable marker gene which can be used to facilitate soybean transformation is a chimeric gene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature 313*:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from *E. coli*; Gritz et al.(1983) *Gene 25*:179-188) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The seed construct comprising the phaseolin 5' region, the fragment encoding the instant polypeptides and the

phaseolin 3' region can be isolated as a restriction fragment. This fragment can then be inserted into a unique restriction site of the vector carrying the marker gene. To 50 µL of a 60 mg/mL 1 µ m gold particle suspension is added (in order): 5 µL DNA (1 µg/ µL), 20 µ L spermidine (0.1 M), and 50 µL $CaCl_2$ (2.5M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400 µL 70% ethanol and resuspended in 40 µ L of anhydrous ethanol. The DNA/particle suspension can be sonicated three times for one second each. Five µ L of the DNA-coated gold particles are then loaded on each macro carrier disk. Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue are normally bombarded. Membrane rupture pressure is set at 1100 psi and the chamber is evacuated to a vacuum of 28 inches of mercury. The tissue is placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post bombardment with fresh media containing 50 mg/mL hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

EXAMPLE 13

Expression of Chimeric Genes in Microbial Cells

The cDNAs encoding the instant polypeptides can be inserted into the T7 *E. coli* expression vector pBT430. This vector is a derivative of pET-3a (Rosenberg et al. (1987) *Gene 56*:125-135) which employs the bacteriophage T7 RNA

polymerase/T7 promoter system. Plasmid pBT430 was constructed by first destroying the EcoR I and Hind III sites in pET-3a at their original positions. An oligonucleotide adaptor containing EcoR I and Hind III sites was inserted at the BamH I site of pET-3a. This created pET-3aM with additional unique cloning sites for insertion of genes into the expression vector. Then, the Nde I site at the position of translation initiation was converted to an Nco I site using oligonucleotide-directed mutagenesis. The DNA sequence of pET-3aM in this region, 5'-CATATGG, was converted to 5'-CCCATGG in pBT430.

Plasmid DNA containing a cDNA may be appropriately digested to release a nucleic acid fragment encoding the protein. This fragment may then be purified on a 1% NuSieve GTG ™ low melting agarose gel (FMC). Buffer and agarose contain 10 μ g/mL ethidium bromide for visualization of the DNA fragment. The fragment can then be purified from the agarose gel by digestion with GELase™ (Epicentre Technologies) according to the manufacturer's instructions, ethanol precipitated, dried and resuspended in 20 μ L of water. Appropriate oligonucleotide adapters may be ligated to the fragment using T4 DNA ligase (New England Biolabs, Beverly, MA). The fragment containing the ligated adapters can be purified from the excess adapters using low melting agarose as described above. The vector pBT430 is digested, dephosphorylated with alkaline phosphatase (NEB) and deproteinized with phenol/chloroform as described above. The prepared vector pBT430 and fragment can then be ligated at 16°C for 15 hours followed by transformation into DH5 electrocompetent cells (GIBCO BRL). Transformants can be selected on agar plates containing LB media and 100 μ g/mL ampicillin. Transformants containing the gene encoding the instant polypeptides are then screened for the correct orientation with respect to the T7 promoter by restriction enzyme analysis.

For high level expression, a plasmid clone with the cDNA insert in the correct orientation relative to the T7 promoter can be transformed into *E. coli* strain BL21(DE3) (Studier et al. (1986) *J. Mol. Biol. 189*:113-130). Cultures are grown in LB medium containing ampicillin (100 mg/L) at 25°C. At an optical density at 600 nm of approximately 1, IPTG (isopropylthio- β -galactoside, the inducer) can be added to a final concentration of 0.4 mM and incubation can be continued for 3 h at 25°C. Cells are then harvested by centrifugation and re-suspended in 50 μL of 50 mM Tris-HCl at pH 8.0 containing 0.1 mM DTT and 0.2 mM phenyl

methylsulfonyl fluoride. A small amount of 1 mm glass beads can be added and the mixture sonicated 3 times for about 5 seconds each time with a microprobe sonicator. The mixture is centrifuged and the protein concentration of the supernatant determined. One µg of protein from the soluble fraction of the culture can be separated by SDS-polyacrylamide gel electrophoresis. Gels can be observed for protein bands migrating at the expected molecular weight.

EXAMPLE 14

Transformation of Somatic Soybean Embryo Cultures

Generic stable soybean transformation protocol:

Soybean embryogenic suspension cultures are maintained in 35 ml liquid media (SB55 or SBP6) on a rotary shaker, 150 rpm, at 28°C with mixed fluorescent and incandescent lights on a 16:8 h day/night schedule. Cultures are subcultured every four weeks by inoculating approximately 35 mg of tissue into 35 ml of liquid medium.

TABLE 17

Stock Solutions (g/L):		**SB55 (per Liter, pH 5.7)**
MS Sulfate 100X Stock		10 ml each MS stocks
$MgSO_4$ $7H_2O$	37.0	1 ml B5 Vitamin stock
$MnSO_4$ H_2O 1.69		0.8 g NH_4NO_3
$ZnSO_4$ $7H_2O$	0.86	3.033 g KNO_3
$CuSO_4$ $5H_2O$	0.0025	1 ml 2,4-D (10mg/mL stock)
MS Halides 100X Stock		60 g sucrose
$CaCl_2$ $2H_2O$	44.0	0.667 g asparagine
KI	0.083	SBP6
$CoCl_2$ $6H_20$	0.00125	same as SB55 except 0.5 ml 2,4-D
KH_2PO_4	17.0	**SB103 (per Liter, pH 5.7)**
H_3BO_3	0.62	1X MS Salts
Na_2MoO_4 $2H_2O$	0.025	6% maltose
MS FeEDTA 100X Stock		750 mg $MgCl_2$
Na_2EDTA	3.724	0.2% Gelrite
$FeSO_4$ $7H_2O$	2.784	**SB71-1 (per Liter, pH 5.7)**
B5 Vitamin Stock		1X B5 salts
10 g m-inositol		1 ml B5 vitamin stock
100 mg nicotinic acid		3% sucrose
100 mg pyridoxine HCl		750 mg $MgCl_2$
1 g thiamine		0.2% Gelrite

Soybean embryogenic suspension cultures are transformed with plasmid DNA by the method of particle gun bombardment (Klein et al (1987) *Nature* 327:70). A DuPont Biolistic PDS1000/HE instrument (helium retrofit) is used for these transformations.

To 50 ml of a 60 mg/ml 1 µm gold particle suspension is added (in order); 5µL DNA(1 µg/ µl), 20 µl spermidine (0.1 M), and 50 µl $CaCl_2$ (2.5 M). The particle preparation is agitated for 3 min, spun in a microfuge for 10 sec and the supernatant removed. The DNA-coated particles are then washed once in 400 µl 70% ethanol and re suspended in 40 µ l of anhydrous ethanol. The DNA/particle suspension is

sonicated three times for 1 sec each. Five µl of the DNA-coated gold particles are then loaded on each macro carrier disk. For selection, a plasmid conferring resistance to hygromycin phosphotransferase (HPT) may be co-bombarded with the silencing construct of interest.

Approximately 300-400 mg of a four week old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue are normally bombarded. Membrane rupture pressure is set at 1000 psi and the chamber is evacuated to a vacuum of 28 inches of mercury. The tissue is placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue is placed back into liquid and cultured as described above.

Eleven days post bombardment, the liquid media is exchanged with fresh SB55 containing 50 mg/ml hygromycin. The selective media is refreshed weekly. Seven weeks post bombardment, green, transformed tissue is observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Thus each new line is treated as an independent transformation event. These suspensions can then be maintained as suspensions of embryos maintained in an immature developmental stage or regenerated into whole plants by maturation and germination of individual somatic embryos.

Independent lines of transformed embryogenic clusters are removed from liquid culture and placed on a solid agar media (SB103) containing no hormones or antibiotics. Embryos are cultured for four weeks at 26°C with mixed fluorescent and incandescent lights on a 16:8 h day/night schedule. During this period, individual embryos are removed from the clusters and screened for alterations in gene expression.

It should be noted that any detectable phenotype, resulting from the co-suppression of a target gene, can be screened at this stage. This would include, but not be limited to, alterations in oil content, protein content, carbohydrate content, growth rate, viability, or the ability to develop normally into a soybean plant.

EXAMPLE 15

Plasmid DNAs for "Complementary Region" Co-suppression

The plasmids in the following experiments are made using standard cloning methods well known to those skilled in the art (Sambrook et al (1989) *Molecular Cloning*, CSHL Press, New York). A starting plasmid pKS18HH (U.S. Patent No. 5,846,784 the contents of which are hereby incorporated by reference) contains a hygromycin B phosphotransferase (HPT) obtained from *E. coli* strain W677 under the control of a T7 promoter and the 35S cauliflower mosaic virus promoter. Plasmid pKS18HH thus contains the T7 promoter/HPT/T7 terminator cassette for expression of the HPT enzyme in certain strains of *E. coli*, such as NovaBlue(DE3) [from Novagen], that are lysogenic for lambda DE3 (which carries the T7 RNA Polymerase gene under lacV5 control). Plasmid pKS18HH also contains the 35S/HPT/NOS cassette for constitutive expression of the HPT enzyme in plants, such as soybean. These two expression systems allow selection for growth in the presence of hygromycin to be used as a means of identifying cells that contain the plasmid in both bacterial and plant systems. pKS18HH also contains three unique restriction endonuclease sites suitable for the cloning other chimeric genes into this vector. Plasmid ZBL100 (PCT Application No. WO 00/11176 published on March 2, 2000) is a derivative of pKS18HH with a reduced NOS 3' terminator. Plasmid pKS67 is a ZBL100 derivative with the insertion of a beta-conglycinin promoter, in front of a NotI cloning site, followed by a phaseolin 3' terminator (described in PCT Application No. WO 94/11516, published on May 26, 1994).

The 2.5 kb plasmid pKS17 contains pSP72 (obtained from Promega Biosystems) and the T7 promoter/HPT/T7 3' terminator region, and is the original vector into which the 3.2 kb BamHI-SalI fragment containing the 35S/HPT/NOS cassette was cloned to form pKS18HH. The plasmid pKS102 is a pKS17 derivative that is digested with XhoI and SalI, treated with mung-bean nuclease to generate blunt ends, and ligated to insert the following linker:

GGCGCGCCAAGCTTGGATCCGTCGACGGCGCGCC SEQ ID NO:97

The plasmid pKS83 has the 2.3 kb BamHI fragment of ML70 containing the Kti3 promoter/NotI/Kti3 3' terminator region (described in PCT Application No. WO 94/11516, published on May 26, 1994) ligated into the BamHI site of pKS17. Additional methods for suppression of endogenous genes are well know in

the art and have been described in the detailed description of the instant invention and can be used to reduce the expression of endogenous cytosolic PPiase gene expression, protein or enzyme activity in a plant cell.

EXAMPLE 16

Suppression by ELVISLIVES Complementary Region

Constructs can be made which have "synthetic complementary regions" (SCR). In this example the target sequence is placed between complementary sequences that are not known to be part of any biologically derived gene or genome (i.e. sequences that are "synthetic" or conjured up from the mind of the inventor). The target DNA would therefore be in the sense or antisense orientation and the complementary RNA would be unrelated to any known nucleic acid sequence. It is possible to design a standard "suppression vector" into which pieces of any target gene for suppression could be dropped. The plasmids pKS106, pKS124, and pKS133 (SEQ ID NO:98) exemplify this. One skilled in the art will appreciate that all of the plasmid vectors contain antibiotic selection genes such as, but not limited to, hygromycin phosphotransferase with promoters such as the T7 inducible promoter.

pKS106 uses the beta-conglycinin promoter while the pKS124 and pKS133 plasmids use the Kti promoter, both of these promoters exhibit strong tissue specific expression in the seeds of soybean. pKS106 uses a 3' termination region from the phaseolin gene, and pKS124 and pKS133 use a Kti 3' termination region. pKS106 and pKS124 have single copies of the 36 nucleotide EagI-ELVISLIVES sequence surrounding a NotI site (the amino acids given in parentheses are back-translated from the complementary strand): SEQ ID NO:99

EagI E L V I S L I V E S NotI

CGGCCG GAG CTG GTC ATC TCG CTC ATC GTC GAG TCG GCGGCCGC

(S) (E) (V) (I) (L) (S) (I) (V) (L) (E) EagI

CGA CTC GAC GAT GAG CGA GAT GAC CAG CTC CGGCCG

pKS133 has 2X copies of ELVISLIVES surrounding the NotI site: SEQ ID NO:100

EagI E L V I S L I V E S EagI E L V I S

cggccggagctggtcatctcgctcatcgtcgagtcg gcggccg gagctggtcatctcg

L I V E S NotI (S)(E (V)(I)(L)(S)(I)(V)(L)(E) EagI
ctcatcgtcgagtcg gcggccgc cgactcgacgatgagcgagatgaccagctc cggccgc

(S)(E)(V)(I)(L)(S)(I)(V)(L)(E) EagI
cgactcgacgatgagcgagatgaccagctc cggccg

The idea is that the single EL linker (SCR) can be duplicated to increase stem lengths in increments of approximately 40 nucleotides. A series of vectors will cover the SCR lengths between 40 bp and the 300 bp. Various target gene lengths can also be evaluated. It is believed that certain combinations of target lengths and complementary region lengths will give optimum suppression of the target, however, it is expected that the suppression phenomenon works well over a wide range of sizes and sequences. It is also believed that the lengths and ratios providing optimum suppression may vary somewhat given different target sequences and/or complementary regions.

The plasmid pKS106 is made by putting the EagI fragment of ELVISLIVES (SEQ ID NO:99) into the NotI site of pKS67. The ELVISLIVES fragment is made by PCR using two primers and no other DNA:

SEQ ID NO:101
5'-
AATTCCGGCCGGAGCTGGTCATCTCGCTCATCGTCGAGTCGGCGGCCGCC
GACTCGACGATGAGCGAGATGACCAGCTCCGGCCGGAATTC-3'

SEQ ID NO:102
5'-GAATTCCGGCCGGAG-3'

The product of the PCR reaction is digested with EagI (5'-CGGCCG-3') and then ligated into NotI digested pKS67. The term “ELVISLIVES” and “EL” are used interchangeably herein.

Additional plasmids can be used to test this example and any synthetic sequence, or naturally occurring sequence, can be used in an analogous manner.

EXAMPLE 17

Screening of transgenic lines for alterations in oil, protein, starch and soluble carbohydrate content.

Transgenic lines can be selected from soybean transformed with a suppression plasmid, such as those described in Example 15 and Example 18. Transgenic lines can be screened for down regulation of cytosolic PPiase in soybean, by measuring alteration in oil, starch, protein, soluble carbohydrate and/or seed weight. Compositional analysis including measurements of seed compositional parameters such as protein content and content of soluble carbohydrates of soybean seed derived from transgenic events that show seed-specific down-regulation of cytosolic, soluble pyrophosphatase genes is performed as follows:

Oil content of mature soybean seed or lyophilized soybean somatic embryos can be measured by NMR as described in Example 2.

Non-structural carbohydrate and protein analysis.

Dry soybean seed are ground to a fine powder in a GenoGrinder and subsamples are weighed (to an accuracy of 0.0001g) into 13x100mm glass tubes; the tubes have Teflon® lined screw-cap closures. Three replicates are prepared for each sample tested. Tissue dry weights are calculated by weighing sub-samples before and after drying in a forced air oven for 18h at 105C.

Lipid extraction is performed by adding 2ml aliquots of heptane to each tube. The tubes are vortex mixed and placed into an ultrasonic bath (VWR Scientific Model 750D) filled with water heated to 60C. The samples are sonicated at full-power (~360W) for 15min and were then centrifuged (5min x 1700g). The supernatants are transferred to clean 13x100mm glass tubes and the pellets are extracted 2 more times with heptane (2ml, second extraction, 1 ml third extraction) with the supernatants from each extraction being pooled. After lipid extraction 1ml acetone is added to the pellets and after vortex mixing, to fully disperse the material, they are taken to dryness in a Speedvac.

Non-structural carbohydrate extraction and analysis.

Two ml of 80% ethanol is added to the acetone dried pellets from above. The samples are thoroughly vortex mixed until the plant material was fully dispersed in the solvent prior to sonication at 60C for 15min. After centrifugation, 5min x 1700g, the supernatants are decanted into clean 13x100mm glass tubes. Two more extractions with 80% ethanol are performed and the supernatants from each are pooled. The extracted pellets are suspended in acetone and dried (as above). An

internal standard β-phenyl glucopyranoside (100ul of a 0.5000 +/- 0.0010g/100ml stock) is added to each extract prior to drying in a Speedvac. The extracts are maintained in a desiccator until further analysis.

The acetone dried powders from above were suspended in 0.9ml MOPS (3-N[Morpholino]propane-sulfonic acid; 50mM, 5mM $CaCl_2$, pH 7.0) buffer containing 100U of heat stable α-amylase (from *Bacillus licheniformis*; Sigma A-4551). Samples are placed in a heat block (90C) for 75min and were vortex mixed every 15min. Samples are then allowed to cool to room temperature and 0.6ml acetate buffer (285mM, pH 4.5) containing 5U amyloglucosidase (Roche 110 202 367 001) is added to each. Samples are incubated for 15 –18h at 55C in a water bath fitted with a reciprocating shaker; standards of soluble potato starch (Sigma S-2630) are included to ensure that starch digestion went to completion.

Post-digestion the released carbohydrates are extracted prior to analysis. Absolute ethanol (6ml) is added to each tube and after vortex mixing the samples were sonicated for 15 min at 60C. Samples were centrifuged (5min x 1700g) and the supernatants were decanted into clean 13x100mm glass tubes. The pellets are extracted 2 more times with 3ml of 80% ethanol and the resulting supernatants are pooled. Internal standard (100ul β-phenyl glucopyranoside, as above) is added to each sample prior to drying in a Speedvac.

Sample preparation and analysis

The dried samples from the soluble and starch extractions described above are solubilized in anhydrous pyridine (Sigma-Aldrich P57506) containing 30mg/ml of hydroxylamine HCl (Sigma-Aldrich 159417). Samples are placed on an orbital shaker (300rpm) overnight and are then heated for 1 hr (75C) with vigorous vortex mixing applied every 15 min. After cooling to room temperature 1ml hexamethyldisilazane (Sigma-Aldrich H-4875) and 100ul trifluoroacetic acid (Sigma-Aldrich T-6508) are added. The samples are vortex mixed and the precipitates are allowed to settle prior to transferring the supernatants to GC sample vials.

Samples are analyzed on an Agilent 6890 gas chromatograph fitted with a DB-17MS capillary column (15m x 0.32mm x 0.25um film). Inlet and detector temperatures are both 275C. After injection (2ul, 20:1 split) the initial column temperature (150C) is increased to 180C at a rate 3C/min and then at 25C/min to a final temperature of 320C. The final temperature is maintained for 10min. The

carrier gas is H_2 at a linear velocity of 51cm/sec. Detection is by flame ionization. Data analysis is performed using Agilent ChemStation software. Each sugar is quantified relative to the internal standard and detector responses were applied for each individual carbohydrate (calculated from standards run with each set of samples). Final carbohydrate concentrations are expressed on a tissue dry weight basis.

Protein Analysis

Protein contents are estimated by combustion analysis on a Thermo Finnigan Flash 1112EA combustion analyzer. Samples, 4-8 mg, weighed to an accuracy of 0.001mg on a Mettler-Toledo MX5 micro balance are used for analysis. Protein contents were calculated by multiplying % N, determined by the analyzer, by 6.25. Final protein contents are expressed on a % tissue dry weight basis.

Additionally, the composition of intact single seed and bulk quantities of seed or powders derived from them may be measured by near-infrared analysis. Measurements of moisture, protein and oil content in soy and moisture, protein, oil and starch content in corn can be measured when combined with the appropriate calibrations.

EXAMPLE 18

Screening of transgenic maize lines for alterations in oil, protein, starch and soluble carbohydrate content.

Transgenic maize lines prepared by the method described in Examples 11 can be screened essentially as described in Example 17. Embryo-specific downregulation of PPiase is expected to lead to an increase in seed oil content. In contrast overexpression of PPiase in the endosperm-specific is expected to lead to an increase in seed starch content.

EXAMPLE 19

Seed specifc RNAi of genes encoding soluble, cytosolic pyrophosphophatase genes in soybean

Three plasmid vectors (pKS420, pKS421, and pKS422) for generation of transgenic soybean events that show seed specific down-regulation of cytosolic pyrophosphophatase genes were constructed.

Briefly plasmid DNA of applicants EST clone ses4d.pk0040.g6 corrresponding to Glyma03g33000 (SEQ ID NO:65) was used in two PCR reactions

with either Primers SA5 (SEQ ID NO:103) and SA7 (SEQ ID NO:104) or SA6 (SEQ ID NO:105) and SA5 (Seq ID NO:103). PCR products from both reactions were gel purified and a mixture of 100 ng of each PCR product was used in a third PCR reaction using only the SA5 PCR primer. A PCR product of 0.83 kb was gel purified, digested with NotI and ligated to NotI linearized, dephosphorylated pBSKS+ (Stratagene, USA). Plasmid DNA was isolated from recombinant clones and digested with NotI. The NotI restriction fragment of 0.83 kb was gel purified and cloned in the sense orientation behind the Kti promoter, to DNA of KS126 (PCT Publication No. WO 04/071467) linearized with the restriction enzyme NotI to give pKS420 (SEQ ID NO:106).

Plasmid DNA of applicants EST clone smj1c.pk008.m18f corrresponding to Glyma11g07530 (Seq ID NO:71) was used in two PCR reactions with either Primers SA8 (Seq ID NO:107) and SA10 (Seq ID NO:108) or SA9 (Seq ID NO:109) and SA8 (Seq ID NO:107). PCR products from both reactions were gel purified and a mixture of 100 ng of each PCR product was used in a third PCR reaction using only PCR primer SA8 (Seq ID NO:107). A PCR product of 0.87 kb was gel purified digested with NotI and ligated to NotI linearized, dephosphorylated pBSKS+ (Stratagene, USA). Plasmid DNA was isolated from recombinant clones and digested with NotI. The NotI restriction fragment of 0.87 kb was gel purified and cloned in the sense orientation behind the Kti promoter, to DNA of KS126 (PCT Publication No. WO 04/071467) linearized with the restriction enzyme NotI to give pKS421 (SEQ ID NO:110).

Plasmid DNA of applicants EST clone sls2a.pk008.i20 corrresponding to Glyma13g19500 (Seq ID NO:111) was used in two PCR reactions with either Primers SA11 (Seq ID NO:113) and SA13 (Seq ID NO:114) or SA12 (Seq ID NO:115) and SA11 (Seq ID NO:113). PCR products from both reactions were gel purified and a mixture of 100 ng of each PCR product was used in a third PCR reaction using only SA11 (Seq ID NO:113) as PCR primer. A PCR product of 0.8 kb was gel purified digested with NotI and ligated to NotI linearized, dephosphorylated pBSKS+ (Stratagene, USA). Plasmid DNA was isolated from recombinant clones and digested with NotI. The NotI restriction fragment of 0.898 kb was gel purified and cloned in the sense orientation behind the Kti promoter, to DNA of KS126 (PCT

Publication No. WO 04/071467) linearized with the restriction enzyme NotI to give pKS422 (SEQ ID NO:116).

Plasmid DNA of KS420, KS421 and KS422 can be used to generate transgenic somatic embryos or seed of soybean using hygromycin selection as described in Example 14. Composition of transgenic somatic embryos or soybean seed generated with pKS420, pKS421 or pKS422 or a combination of these plasmids can be determined as described in Example 17.

EXAMPLE 20

Compositional analysis of arabidopsis events transformed with DNA constructs for silencing of cytosolic pyrophosphatase genes

The example describes seed composition of transgenic events gene generated with pKR1482-PPA1. It demonstrates that transformation with DNA constructs for silencing of genes encoding cytosolic pyrophosphatases leads to increased oil content that is accompanied by a reduction of seed storage protein content and (to a smaller extend a reduction) in soluble carbohydrates.

Three transgenic events K44615, K44696 and K44698 were generated by agrobacterium-mediated transformation with pKR1482-PPA1 (SEQ ID NO:28) as described in Example 5.

T3 seed of K44615 and T2 seed of K44696 and K44698 were germinated on selective plant growth media containing kanamycin. Kanamycin-resistant seedlings were transferred to soil and grown alongside untransformed control plant as described in Example 5. At maturity seeds were bulk-harvested from transgenic lines and control plants and subjected to oil analysis by NMR as described in Example 2. The seed sample were subjected to compositional analysis of protein and soluble carbohydrate content of triplicate samples as described in Example 2.

TABLE 18

Seed composition of arabidospis events transformed with DNA constructs for silencing of cytosolic pyrophosphatase genes

Genotype	Bar code ID	Oil (%, NMR)	Protein %	fructose (µg mg^{-1} seed)	glucose (µg mg^{-1} seed)
pKR1482-PPA1 (T4)	K44615	44.3	16.97	0.47	3.21
WT		40.7	18.51	0.41	3.19
	Δ TG/WT %	8.7	-8.3	13.7	0.6
Genotype	Bar code ID	sucrose (µg mg^{-1} seed)	raffinose (µg mg^{-1} seed)	stachyose (µg mg^{-1} seed)	total soluble CHO (µg mg^{-1} seed)
pKR1482-PPA1 (T4)	K44615	15.04	0.44	1.04	20.82
WT		15.34	0.43	1.08	21.06
	Δ TG/WT %	-2.0	1.8	-4.1	-1.1
Genotype	Bar code ID	Oil (%, NMR)	Protein %	fructose (µg mg^{-1} seed)	glucose (µg mg^{-1} seed)
pKR1482-PPA1 (T3)	K44696	44.4	16.00	0.42	3.34
WT		42.2	18.68	0.37	3.51
	Δ TG/WT %	5.2	-14.3	12.4	-4.7
Genotype	Bar code ID	sucrose (µg mg^{-1} seed)	raffinose (µg mg^{-1} seed)	stachyose (µg mg^{-1} seed)	total soluble CHO (µg mg^{-1} seed)
pKR1482-PPA1 (T3)	K44696	15.11	0.42	1.20	21.13
WT		16.23	0.46	1.34	22.4
	Δ TG/WT %	-6.9	-9.3	-10.4	-5.7

Genotype	Bar code ID	Oil (%, NMR)	Protein %	fructose (µg mg^{-1} seed)	glucose (µg mg^{-1} seed)
pKR1482-PPA1 (T3)	K44698	45.4	15.38	0.43	2.98
WT		43.3	17.74	0.41	4.13
	Δ TG/WT %	4.9	-13.3	5.5	-27.8
Genotype	Bar code ID	sucrose (µg mg^{-1} seed)	raffinose (µg mg^{-1} seed)	stachyose (µg mg^{-1} seed)	total soluble CHO (µg mg^{-1} seed)
pKR1482-PPA1 (T3)	K44698	15.18	0.43	1.50	21.04
WT		15.65	0.45	1.56	22.69
	Δ TG/WT %	-3.0	-4.2	-3.9	-7.3

Table 18 demonstrates that the oil increase associated with the presence of the pKR1482-PPA1 transgene (SEQ ID NO:28) is accompanied by a reduction in seed protein content and a small reduction in soluble carbohydrate content. The latter was calculated by summarizing the content of pinitol, sorbitol, fructose, glucose, myo-Inositol, sucrose, raffinose and stachyose.

EXAMPLE 21

Expression of genes from maize and soybean encoding cytosolic pyrophosphatases alters oil content of arabidopsis seed

The example describes the generation of vectors for seed –specific expression of pyrophosphatase genes from soybean and maize in transgenic arabidopsis plants and analysis of seed oil content of related transgenic lines.

Plasmid DNA of applicants EST clone smj1c.pk008.m18 corresponding to Glyma11g07530 (SEQ ID NO: 71) was used in a PCR reaction with primers SA236 (SEQ ID NO:117) and SA237 (SEQ ID NO:118). PCR products were cloned into the ***pCR8*** TOPO TA ***vector (Invitrogen, CA, USA) according to manufacturer instructions.*** Purified plasmid DNA of ***pCR8*** containing the Glyma11g07530 ORF, pKR1478 (SEQ ID NO:9) and LR recombinase ***(Invitrogen, CA, USA)*** were used

according to manufacturer instructions thus generating binary vector pKR1478- Glyma11g07530 which is set forth as SEQ ID NO:119 .

Plasmid DNA of applicants EST clone cds3f.pk005.n3corresponding to PCO640614 (SEQ ID NO: 77) was used in a PCR reaction with primers SA242 (SEQ ID NO:120) and SA243(SEQ ID NO:121). PCR products were cloned into the ***pCR8*** TOPO TA ***vector (Invitrogen, CA, USA) according to manufacturer instructions.*** Purified plasmid DNA of ***pCR8*** containing the PCO640614 ORF, pKR1478 (SEQ ID NO:9) and LR recombinase ***(Invitrogen, CA, USA)*** were used ***according to manufacturer instructions thus generating binary vector*** pKR1478- PCO640614 which is set forth as SEQ ID NO:122 .

Plasmid DNA of applicants EST clone ciec.pk020.010 corresponding to PCO650999 (SEQ ID NO: 81) was used in a PCR reaction with primers SA245 (SEQ ID NO:123) and SA246 (SEQ ID NO:124). PCR products were cloned into the ***pCR8*** TOPO TA ***vector (Invitrogen, CA, USA) according to manufacturer instructions.*** Purified ***pCR8*** plasmid DNA containing PCO650999, pKR1478 (SEQ ID NO:9) and LR recombinase ***(Invitrogen, CA, USA)*** were used ***according to manufacturer instructions thus generating binary vector*** pKR1478- PCO650999 which is set forth as SEQ ID NO:125 .

Plasmid DNA of pKR1478- Glyma11g07530 (SEQ ID NO:119), pKR1478-PCO640614 (SEQ ID NO:122) and pKR1478- PCO650999 (SEQ ID NO:125) were used for agrobacterium-mediated transformation of arabidopsis plant as described in Examle 4. T1 plants representing unique transgenic events were grown alongside WT arabidopsis plants as described previously (Example 4). Seed oil content of T1 and control plants was measured by NMR as described in Example 2 and is listed in Tables 19 –21. In these tables the oil content of a given transgenic event is compared to the average oil content of 8-12 WT control plants grown alongside the transgenic lines.

Tables 19 - 21, show that seed specific expression of genes encoding cytosolic pyrophosphatases from soy and maize leads to reduced oil accumulation in transgenic arabidopsis plants.

TABLE 19

Seed oil content of arabidopsis T1 plants generated with pKR1478- Glyma11g07530

BARCODE	% oil	oil content % of wt	avg. oil content % of WT
K57442	43.1	104.3	
K57444	42.0	101.7	
K57443	41.8	101.2	
K57448	41.0	99.2	
K57439	40.9	99.1	
K57446	40.7	98.6	
K57449	40.7	98.5	
K57447	40.6	98.3	
K57445	40.2	97.2	
K57441	40.0	96.8	
K57440	37.5	90.9	
K57438	36.7	88.7	97.9
WT (avg)	41.3		

TABLE 20

Seed oil content of arabidopsis T1 plants generated with pKR1478- PCO640614

BARCODE	% oil	oil content % of wt	avg. oil content % of WT
K57333	42.9	103.6	
K57334	42.9	103.6	
K57344	42.2	102.0	
K57335	41.7	100.8	
K57339	40.8	98.7	
K57338	40.6	98.1	
K57342	40.5	97.9	
K57343	40.4	97.7	
K57337	40.0	96.7	
K57341	39.4	95.3	
K57345	39.3	95.0	
K57340	37.5	90.7	
K57336	37.3	90.2	
K57346	34.1	82.5	96.6
WT (avg)	41.4		

TABLE 21

Seed oil content of arabidopsis T1 plants generated with pKR1478- PCO650999

BARCODE	% oil	oil content % of wt	avg. oil content % of WT
K57498	44.5	104.2	
K57490	44.1	103.1	
K57510	43.7	102.3	
K57492	43.6	102.0	
K57502	43.4	101.5	
K57508	43.0	100.7	
K57497	42.7	99.9	
K57489	42.6	99.7	
K57500	42.3	98.9	
K57501	42.0	98.2	
K57491	42.0	98.2	
K57493	41.7	97.5	
K57495	41.5	97.2	
K57509	41.3	96.7	
K57496	40.9	95.8	
K57494	40.9	95.7	
K57499	40.9	95.7	
K57505	40.5	94.7	
K57504	39.3	91.9	
K57503	38.1	89.1	
K57506	37.6	88.1	
K57507	35.7	83.5	97.0
WT (avg)	42.7		

CLAIMS

What is claimed is:

1. A transgenic plant comprising a recombinant DNA construct comprising a polynucleotide operably linked to at least one regulatory element, wherein said polynucleotide encodes a polypeptide having an amino acid sequence of at least 70% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 or 112 and wherein seed obtained from said transgenic plant has an altered i.e. increased or decreased oil, protein, starch and/or soluble carbohydrate content when compared to a control plant not comprising said recombinant DNA construct.

2. A transgenic seed obtained from the transgenic plant of claim 1 comprising a recombinant DNA construct comprising a polynucleotide operably linked to at least one regulatory element, wherein said polynucleotide encodes a polypeptide having an amino acid sequence of at least 70% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 or 112 and wherein said transgenic seed has an altered oil, protein, starch and/or soluble carbohydrate content when compared to a seed from a control plant not comprising said recombinant DNA construct.

3. A transgenic seed obtained from the transgenic plant of claim 1 comprising a recombinant DNA construct comprising a polynucleotide operably linked to at least one regulatory element, wherein said polynucleotide encodes a polypeptide having an amino acid sequence of at least 70% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 or 112 and wherein said transgenic seed has an increased starch content of at least 0.5% when compared to a seed from a control plant not comprising said recombinant DNA construct.

4. A transgenic seed comprising:

a recombinant DNA construct comprising:

(a) a polynucleotide operably linked to at least one regulatory element, wherein said polynucleotide encodes a polypeptide having an amino

acid sequence of at least 70% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 or 112 or

(b) a suppression DNA construct comprising at least one regulatory element operably linked to:

(i) all or part of: (A) a nucleic acid sequence encoding a polypeptide having an amino acid sequence of at least 70% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: : 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 or 112, or (B) a full complement of the nucleic acid sequence of (b)(i)(A); or

(ii) a region derived from all or part of a sense strand or antisense strand of a target gene of interest, said region having a nucleic acid sequence of at least 70% sequence identity, based on the Clustal V method of alignment, when compared to said all or part of a sense strand or antisense strand from which said region is derived, and wherein said target gene of interest encodes a cytosolic Pyrophosphatase,

and wherein said plant has an altered oil, protein, starch and/or soluble carbohydrate content when compared to a control plant not comprising said recombinant DNA construct.

5. A transgenic seed having an increased oil content of at least 2% when compared to the oil content of a non-transgenic seed, wherein said transgenic seed comprises a recombinant DNA construct comprising:

(a) all or part of the nucleotide sequence set forth in SEQ ID NO: 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95 or 111;

or (b) the full-length complement of (a):

wherein (a) or (b) is of sufficient length to inhibit expression of endogenous cytosolic pyrophosphatase activity in a transgenic plant and further wherein said seed has an

increase in oil content of at least 2%, on a dry-weight basis, as compared to seed obtained from a non-transgenic plant.

6. A transgenic seed comprising a recombinant DNA construct comprising:

(a) all or part of the nucleotide sequence set forth in SEQ ID NO: 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95 or 111;

or (b) the full-length complement of (a):

wherein (a) or (b) is of sufficient length to inhibit expression of endogenous cytosolic pyrophosphatase activity in a transgenic plant and further wherein said seed has an increase in oil content of at least 2%, on a dry-weight basis, as compared to seed obtained from a non-transgenic plant.

7. A method for producing a transgenic plant, the method comprising:

(a) transforming a plant cell with a recombinant DNA construct comprising a polynucleotide operably linked to at least one regulatory sequence, wherein the polynucleotide encodes a polypeptide having an amino acid sequence of at least 70% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 or 112; and

(b) regenerating a plant from the transformed plant cell.

8. A method for producing transgenic seeds, the method comprising:

(a) transforming a plant cell with a recombinant DNA construct comprising a polynucleotide operably linked to at least one regulatory sequence, wherein the polynucleotide encodes a polypeptide having an amino acid sequence of at least 70% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 or 112; and

(b) regenerating a transgenic plant from the transformed plant cell of (a); and

(c) selecting a transgenic plant that produces a transgenic seed having an altered oil, protein, starch and/or soluble carbohydrate content, as compared to a seed obtained from a non-transgenic plant.

9. The transgenic seed of any one of claim 1, 2, 3, 4, 5 or 6, wherein the transgenic seed is obtained from a monocot or dicot plant.

10. The transgenic seed of any one of claim 1, 2, 3, 4, 5 or 6, wherein the at least one regulatory element is a seed-specific or seed-preferred promoter.

11. The method of any one of claims 7 or 8, wherein the transgenic seed is obtained from a transgenic soybean plant comprising in its genome the recombinant construct.

12. A product and/or by-product obtained from the transgenic seed of claim 10.

13. The transgenic seed obtained by the method of claim 7 or 8, wherein the transgenic seed is obtained from a monocot or dicot plant.

14. A product and/or by-product from transgenic seed of claim 2, wherein the plant is maize or soybean.

15. A product and/or by-product from the transgenic seed of claim 3, wherein the plant is maize or soybean.

16. A product and/or by-product from the transgenic seed of claim 4, wherein the plant is maize or soybean.

17. A product and/or by-product from the transgenic seed of claim 5, wherein the plant is maize or soybean.

18. A product and/or by-product from the transgenic seed of claim 6, wherein the plant is maize or soybean.

FIG. 1A

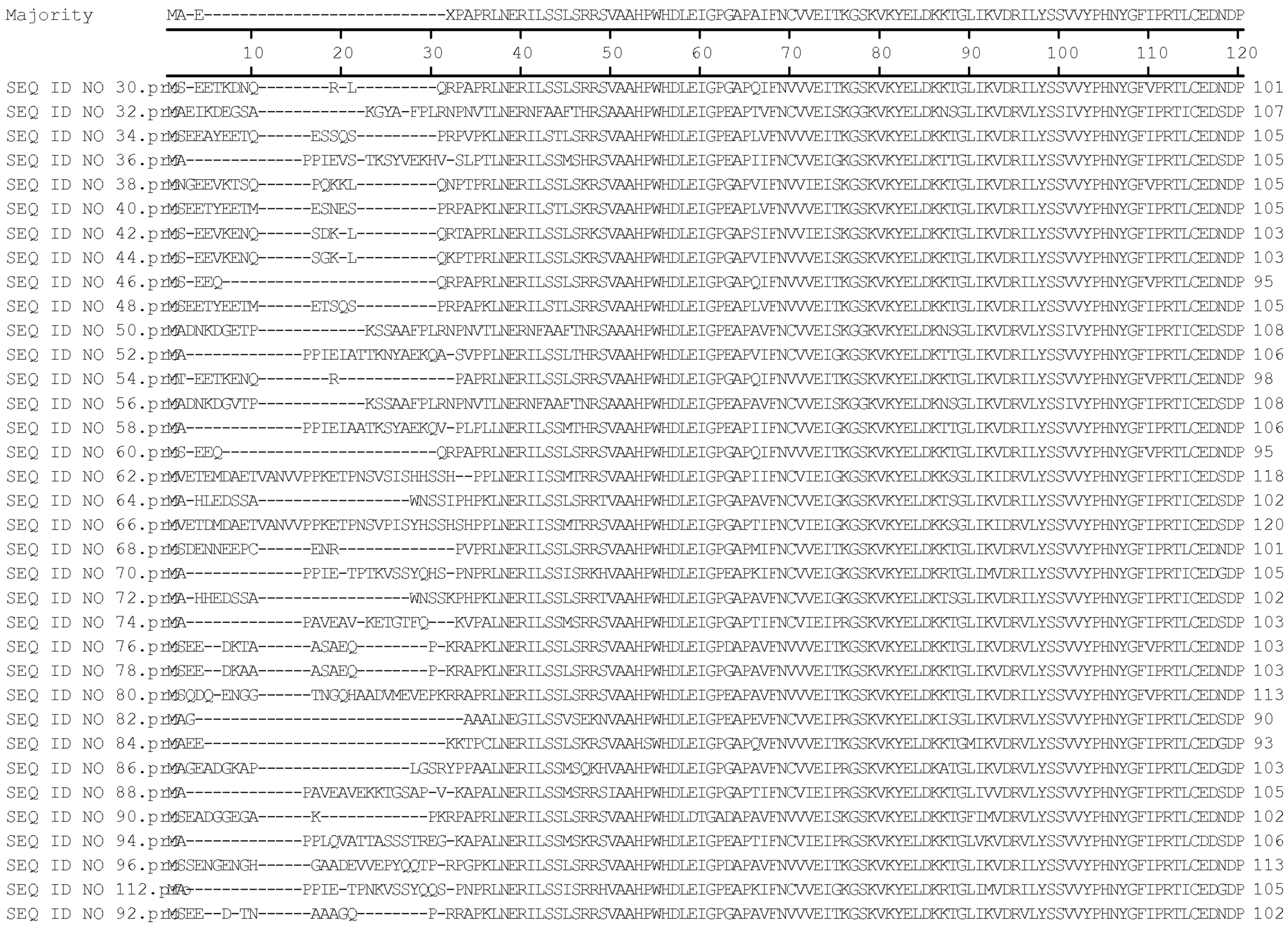

Majority MA-E----------XPAPRLNERILSSLSRRSVAAHPWHDLEIGPGAPAIFNCVVEITKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFIPRTLCEDNDP
10 20 30 40 50 60 70 80 90 100 110 120
SEQ ID NO 30.pro MS-EETKDNQ--------R-L---------QRPAPRLNERILSSLSRRSVAAHPWHDLEIGPGAPQIFNVVVEITKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFVPRTLCEDNDP 101
SEQ ID NO 32.pro MAEIKDEGSA------------KGYA-FPLRNPNVTLNERNFAAFTHRSAAAHPWHDLEIGPFAPTVFNCVVEISKGGKVKYELDKNSGLIKVDRVLYSSIVYPHNYGFIPRTICEDSDP 107
SEQ ID NO 34.pro MSEEAYEETQ------ESSQS---------PRPVPKLNERILSTLSRRSVAAHPWHDLEIGPFAPLVFNVVVEITKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFIPRTLCEDNDP 105
SEQ ID NO 36.pro MA------------PPIEVS-TKSYVEKHV-SLPTLNERILSSMSHRSVAAHPWHDLEIGPFAPIIFNCVVEIGKGSKVKYELDKTTGLIKVDRILYSSVVYPHNYGFIPRTLCEDSDP 105
SEQ ID NO 38.pro MNGEEVKTSQ------PQKKL---------QNPTPRLNERILSSLSKRSVAAHPWHDLEIGPGAPVIFNVVIEISKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFVPRTLCEDNDP 105
SEQ ID NO 40.pro MSEETYEETM------ESNES---------PRPAPKLNERILSTLSKRSVAAHPWHDLEIGPFAPLVFNVVVEITKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFIPRTLCEDNDP 105
SEQ ID NO 42.pro MS-EEVKENQ------SDK-L---------QRTAPRLNERILSSLSRKSVAAHPWHDLEIGPGAPSIFNVVIEISKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFIPRTLCEDNDP 103
SEQ ID NO 44.pro MS-EEVKENQ------SGK-L---------QKPTPRLNERILSSLSKRSVAAHPWHDLEIGPGAPVIFNVVVEISKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFIPRTLCEDNDP 103
SEQ ID NO 46.pro MS-EEQ----------QRPAPRLNERILSSLSRRSVAAHPWHDLEIGPGAPQIFNVVVEITKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFVPRTLCEDNDP 95
SEQ ID NO 48.pro MSEETYEETM------ETSQS---------PRPAPKLNERILSTLSRRSVAAHPWHDLEIGPFAPLVFNVVVEITKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFIPRTLCEDNDP 105
SEQ ID NO 50.pro MADNKDGETP----------KSSAAFPLRNPNVTLNERNFAAFTNRSAAAHPWHDLEIGPFAPAVFNCVVEISKGGKVKYELDKNSGLIKVDRVLYSSIVYPHNYGFIPRTICEDSDP 108
SEQ ID NO 52.pro MA------------PPIEIATTKNYAEKQA-SVPPLNERILSSLTHRSVAAHPWHDLEIGPFAPVIFNCVVEIGKGSKVKYELDKTTGLIKVDRILYSSVVYPHNYGFIPRTLCEDNDP 106
SEQ ID NO 54.pro MT-EETKENQ--------R----------PAPRLNERILSSLSRRSVAAHPWHDLEIGPGAPQIFNVVVEITKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFVPRTLCEDNDP 98
SEQ ID NO 56.pro MADNKDGVTP----------KSSAAFPLRNPNVTLNERNFAAFTNRSAAAHPWHDLEIGPFAPAVFNCVVEISKGGKVKYELDKNSGLIKVDRVLYSSIVYPHNYGFIPRTICEDSDP 108
SEQ ID NO 58.pro MA------------PPIEIAATKSYAEKQV-PLPLLNERILSSMTHRSVAAHPWHDLEIGPFAPIIFNCVVEIGKGSKVKYELDKTTGLIKVDRILYSSVVYPHNYGFIPRTLCEDNDP 106
SEQ ID NO 60.pro MS-EEQ----------QRPAPRLNERILSSLSRRSVAAHPWHDLEIGPGAPQIFNVVVEITKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFVPRTLCEDNDP 95
SEQ ID NO 62.pro MVETEMDAETVANVVPPKETPNSVSISHHSSH--PPLNERIISSMIRRSVAAHPWHDLEIGPGAPIIFNCVIEIGKGSKVKYELDKKSGLIKIDRVLYSSVVYPHNYGFIPRTICEDSDP 118
SEQ ID NO 64.pro MA-HLEDSSA----------WNSSIPHPKLNERILSSLSRRTVAAHPWHDLEIGPGAPAVFNCVVEIGKGSKVKYELDKTSGLIKVDRILYSSVVYPHNYGFIPRTICEDSDP 102
SEQ ID NO 66.pro MVETDMDAETVANVVPPKETPNSVPISYHSSHSHPPLNERIISSMIRRSVAAHPWHDLEIGPGAPTIFNCVIEIGKGSKVKYELDKKSGLIKIDRVLYSSVVYPHNYGFIPRTICEDSDP 120
SEQ ID NO 68.pro MSDENNEEPC------ENR----------PVPRLNERILSSLSRRSVAAHPWHDLEIGPGAPMIFNCVVEITKGSKVKYELDKKTGLIKVDRVLYSSVVYPHNYGFIPRTLCEDNDP 101
SEQ ID NO 70.pro MA----------PPIE-TPTKVSSYQHS-PNPRLNERILSSISRKHVAAHPWHDLEIGPFAPKIFNCVVEIGKGSKVKYELDKRTGLIMVDRILYSSVVYPHNYGFIPRTICEDGDP 105
SEQ ID NO 72.pro MA-HHEDSSA----------WNSSKPHPKLNERILSSLSRRTVAAHPWHDLEIGPGAPAVFNCVVEIGKGSKVKYELDKTSGLIKVDRILYSSVVYPHNYGFIPRTICEDSDP 102
SEQ ID NO 74.pro MA----------PAVEAV-KETGTFQ---KVPALNERILSSMSRRSVAAHPWHDLEIGPGAPTIFNCVIEIPRGSKVKYELDKKTGLIKVDRVLYSSVVYPHNYGFIPRTLCEDSDP 103
SEQ ID NO 76.pro MSEE--DKTA------ASAEQ--------P-KRAPKLNERILSSLSRRSVAAHPWHDLEIGPDAPAVFNVVVEITKGSKVKYELDKKTGLIKVDRVLYSSVVYPHNYGFVPRTLCEDNDP 103
SEQ ID NO 78.pro MSEE--DKAA------ASAEQ--------P-KRAPKLNERILSSLSRRSVAAHPWHDLEIGPGAPAVFNVVVEITKGSKVKYELDKKTGLIKVDRVLYSSVVYPHNYGFIPRTLCEDNDP 103
SEQ ID NO 80.pro MSQDQ-ENGG------TNGQHAADVMEVEPKRRAPRLNERILSSLSRRSVAAHPWHDLEIGPFAPAVFNVVVEITKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFVPRTLCEDNDP 113
SEQ ID NO 82.pro MAG----------AAALNEGILSSVSEKNVAAHPWHDLEIGPFAPEVFNCVVEIPRGSKVKYELDKISGLIKVDRVLYSSVVYPHNYGFIPRTLCEDSDP 90
SEQ ID NO 84.pro MAEE----------KKTPCLNERILSSLSKRSVAAHSWHDLEIGPGAPQVFNVVVEITKGSKVKYELDKKTGMIKVDRVLYSSVVYPHNYGFIPRTLCEDGDP 93
SEQ ID NO 86.pro MAGEADGKAP----------LGSRYPPAALNERILSSMSQKHVAAHPWHDLEIGPGAPAVFNCVVEIPRGSKVKYELDKATGLIKVDRVLYSSVVYPHNYGFIPRTLCEDGDP 103
SEQ ID NO 88.pro MA----------PAVEAVEKKTGSAP-V-KAPALNERILSSMSRRSIAAHPWHDLEIGPGAPTIFNCVIEIPRGSKVKYELDKKTGLIVVDRVLYSSVVYPHNYGFIPRTLCEDSDP 105
SEQ ID NO 90.pro MSEADGGEGA------K----------PKRPAPRLNERILSSLSRRSVAAHPWHDLDTGADAPAVFNVVVEISKGSKVKYELDKKTGFIMVDRVLYSSVVYPHNYGFIPRTLCEDNDP 102
SEQ ID NO 94.pro MA----------PPLQVATTASSSTREG-KAPALNERILSSMSKRSVAAHPWHDLEIGPFAPTIFNCVIEIPRGSKVKYELDKKTGLVKVDRVLYSSVVYPHNYGFIPRTLCDDSDP 106
SEQ ID NO 96.pro MSSENGENGH------GAADEVVEPYQQTP-RPGPKLNERILSSLSRRSVAAHPWHDLEIGPDAPAVFNVVVEITKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFIPRTLCEDNDP 113
SEQ ID NO 112.pro MAG----------PPIE-TPNKVSSYQQS-PNPRLNERILSSISRRHVAAHPWHDLEIGPFAPKIFNCVVEIGKGSKVKYELDKRTGLIMVDRILYSSVVYPHNYGFIPRTICEDGDP 105
SEQ ID NO 92.pro MSEE--D-TN------AAAGQ--------P-RRAPKLNERILSSLSRRSVAAHPWHDLEIGPGAPAVFNVVVEITKGSKVKYELDKKTGLIKVDRVLYSSVVYPHNYGFIPRTLCEDNDP 102

FIG.1B

```
Majority          MDVLVLMQEPVLPGCFLRARAIGLMPMIDQGEKDDKIIAVCADDPEYRHYNDIKELPPHRLAEIRRFFEDYKKNENKEVAVNDFLPAEAAIEAIQYSMDLYAEYILETLRR-
                                  130       140       150       160       170       180       190       200       210       220       230

SEQ ID NO 30.pro  IDVLVIMQEPVLPGCFLRARAIGLMPMIDQGEKDDKIIAVCVDDPEYKHYTDIKELPPHRLSEIRRFFEDYKKNENKEVAVNDFLPSESAVEAIQYSMDLYAEYILHTLRR.  213
SEQ ID NO 32.pro  MDVLVLMQEPVLTGSFLRARAIGLMPMIDQGEKDDKIIAVCADDPEFRHYRDIKELPPHRLAEIRRFFEDYKKNENKKVDVEAFLPAQAAIDAIKDSMDLYAAYIKAGLQR.  219
SEQ ID NO 34.pro  LDVLVLMQEPVLPGCFLRARAIGLMPMIDQGEKDDKIIAVCADDPEYKHFTDIKQLAPHRLQEIRRFFEDYKKNENKKVAVNDFLPSESAHEAIQYSMDLYAEYILHTLRR.  217
SEQ ID NO 36.pro  IDVLVIMQEPVIPGCFLRAKAIGLMPMIDQGEKDDKIIAVCADDPEYRHYNDISELPPHRMAEIRRFFEDYKKNENKEVAVNDFLPATAAYDAVQHSMDLYADYVVENLRR.  217
SEQ ID NO 38.pro  IDVLVIMQEPVLPGCFLRARAIGLMPMIDQGEKDDKIIAVCVDDPEYKHITNINELPPHRLSEIRRFFEDYKKNENKEVAVNDFLQPGPAIEAIQYSMDLYAEYILHTLRR.  217
SEQ ID NO 40.pro  LDVLVLMQEPVLPGCFLRARAIGLMPMIDQGEMDDKIIAVCADDPEYKHFTDIKQLAPHRLSEIRRFFEDYKKNEHKEVAVNDFLPSEKAHEAIQYSMDLYAEYILHSLRR.  217
SEQ ID NO 42.pro  LDVLVIMQEPVLPGCFLRARAIGLMPMIDQGEKDDKIIAVCVDDPEYKHYTDIKELPPHRLSEIRRFFEDYKKNENKEVAVNDFLPNGPAVEAIQYSMDLYAEYILHTLRR.  215
SEQ ID NO 44.pro  LDVLVIMQEPVLPGCFLRARAIGLMPMIDQGEKDDKIIAVCVDDPEYKHYTDIKELPPHRLTEIRRFFEDYKKNENKEVAVNDFLPNGPAVEAIQYSMDLYAEYILHTLRR.  215
SEQ ID NO 46.pro  IDVLVIMQEPVLPGCFLRARAIGLMPMIDQGEKDDKIIAVCVDDPEYKHYIDIKELPPHRLSEIRRFFEDYKKNENKEVAVNDFLPSENAIDAIQYSMDLYAEYILHTLRR.  207
SEQ ID NO 48.pro  LDVLVLMQEPVLPGCFLRARAIGLMPMIDQGEMDDKIIAVCADDPEYKHFTDIKQLAPHRLSEIRRFFEDYKKNEHKEVAVNDFLPSEKAHEAIQYSMDLYAEYILHTLRR.  217
SEQ ID NO 50.pro  IDVLVLMQEPVLTGSFLRARAIGLMPMIDQGEKDDKIIAVCADDPEFRHYRDIKELPPHRLAEIRRFFEDYKKNENKKVAVEGFLPAQAAIDAIKDSMDLYAAYIKAGLQR.  220
SEQ ID NO 52.pro  IDVLVIMQEPVIPGCFLRAKAIGLMPMIDQGEKDDKIIAVCADDPEYRHYNDIKELPPHRLAEIRRFFEDYKKNENKEVAVNDFLPATAAYEAVQHSMDLYADYVMETLRR.  218
SEQ ID NO 54.pro  IDVLVIMQEPVLPGCFLRARAIGLMPMIDQGEKDDKIIAVCVDDPEYKHYTDIKELPPHRLSEIRRFFEDYKKNENKEVAVNDFLPSEKAIEAIQYSMDLYAEYILHTLRR.  210
SEQ ID NO 56.pro  IDVLVLMQEPVLTGSFLRARAIGLMPMIDQGEKDDKIIAVCADDPEFRHYRDIKELPPHRLAEIRRFFEDYKKNENKKVAVEGFLPAQAAIDAIKDSMDLYAAYIKAGLQR.  220
SEQ ID NO 58.pro  IDVLVIMQEPVIPGCFLRAKAIGLMPMIDQGEKDDKIIAVCADDPEYRHYNDIKELPPHRLAEIRRFFEDYKKNENKEVAVNDFLPATAAYEAVQHSMDLYADYVMETLRR.  218
SEQ ID NO 60.pro  IDVLVIMQEPVLPGCFLRARAIGLMPMIDQGEKDDKIIAVCVDDPEYKHYIDIKELPPHRLSEIRRFFEDYKKNENKEVAVNDFLPSENAIDAIQYSMDLYAEYILHTLRR.  207
SEQ ID NO 62.pro  LDVLIIMQEPVLPGCFLRAKAIGLMPMIDQGEKDDKIIAVCADDPEYRHYNDIKELPPHRLAEIRRFFEDYKKNENKEVAVNDFLPASAAFEAVKRSMSLYADYIVESLRR.  230
SEQ ID NO 64.pro  MDVLVLMQEPVLPGSFLRARAIGLMPMIDQGERDDKIIAVCADDPEFRHYTDINELPPHRLAEIRRFFEDYKKNENKIVDVEDFLPAEAAIDAINYSMDLYAAYIVESLRH.  214
SEQ ID NO 66.pro  LDVLIIMQEPVLPGCFLRAKAIGLMPMIDQGEKDDKIIAVCADDPEYRHYNDIKELPPHRLAEIRRFFEDYKKNENKEVAVNDFLPASAAFEAVNRSMSLYADYIVESLRR.  232
SEQ ID NO 68.pro  IDVLVLMQEPVLPGCFLRARAIGLMPMIDQGEKDDKIIAVCADDPEYKHYTDFKELPPHRLMEIRRFFEDYKKNENKEVAVNDFLPASTAVESIQYSMDLYAEYILHTLRR.  213
SEQ ID NO 70.pro  MDVLVIMQEPVLPGCFLRAKAIGLMPMIDQGEKDDKIIAVCADDPEYRHYNDIKDLPPHRLAEIRRFFEDYKKNENKEVAVNDFLPASAAYEAIKHSMTLYAEYVVENLRR.  217
SEQ ID NO 72.pro  MDVLVLMQEPVLPGSFLRARAIGLMPMIDQGERDDKIIAVCADDPEFRHYTDIKELPPHRLAEIRRFFEDYKKNENKIVDVEDFLPAEAAIDAIKYSMDLYAAYIVESLRH.  214
SEQ ID NO 74.pro  LDVLVIMQEPVIPGCFLRAKAIGLMPMIDQGEADDKIIAVCADDPEYRHYNDIKELPPHRLAEIRRFFEDYKKNENKEVAVNDFLPASAAYEAIQHSMDLYATYIVEGLRR.  215
SEQ ID NO 76.pro  MDVLVLMQEPVVPGSFLRARAIGLMPMIDQGEKDDKIIAVCADDPEYRHYNDISELSPHRLQEIKRFFEDYKKNENKEVAVDAFLPATTAREAIQYSMDLYAQYILQSLRQ.  215
SEQ ID NO 78.pro  MDVLVLMQEPVVPGSFLRARAIGLMPMIDQGEKDDKIIAVCADDPEYRHYNDISELSPHRLQEIKRFFEDYKKNENKEVAVDAFLPATTAQEAIQYSMDLYAQYILQSLRQ.  215
SEQ ID NO 80.pro  MDVLVLMQEPVLPGAFLRARAIGLMPMIDQGEKDDKIIAVCADDPEYRHYNDISELSPHRLQEIRRFFEDYKKNENKEVAVNDFLPAAAAREAIQYSMDLYGQYIMQTLRR.  225
SEQ ID NO 82.pro  MDVLVLMQEQVVPGCFLRARAIGLMPMIDQGEKDDKIIAVCADDPEFRHYKDISDLPPHRLQEIRRFFEDYKKNENKEVAVNDFLPAEDAIKAIEHSMDLYGSYVMESLRR.  202
SEQ ID NO 84.pro  MDVLVLMQEPVIPGCYLRAKAIGLMPMIDQGEKDDKIIAVCVDDPEFRHFNDLKELSPHRLAEIRRFFEDYKKNENKEVAVNDFLPPATAQEAIKYSMDLYAEYILHSLRR.  205
SEQ ID NO 86.pro  MDVLVLMQEQVVPGCFLRARAIGLMPMIDQGEKDDKIIAVCADDPEYRHFRDIKEIPPHRLQEIRRFFEDYKKNENKEVAVNEFLPAEDAINAIKYSMDLYGAYIIESLRK.  215
SEQ ID NO 88.pro  LDVLVIMQEPVIPGCFLRAKAIGLMPMIDQGEADDKIIAVCADDPEYKHYNDIKELPPHRLAEIRRFFEDYKKNENKEVAVNDFLPASAAYEAIKHSMDLYATYIVEGLRR.  217
SEQ ID NO 90.pro  MDVLVLMQEPVIPGCFLRARAIGLMPMIDQGEKDDKIIAVCVDDPEYRHYNDLSELSPHRVQEIRRFFEDYKKNENKEVAVNEVLPVTAARDAIQYSMDLYAQYIEH-LGQ.  213
SEQ ID NO 94.pro  LDVLVIMQEPVIPGCFLRAKAIGVMPMIDQGEADDKIIAVCADDPEYKHYNDIKDLPPHRLAEIRRFFEDYKKNENKEVAVNDFMPATSAYETIRHSMDLYATYILEGLRR.  218
SEQ ID NO 96.pro  MDVLVLMQEPVLPGSFLRARAIGLMPMIDQGEKDDKIIAVCADDPEYRHFNNLSELSPHRLQEIRRFFEDYKKNENKEVAVNDFLPAPTAREAIQYSMDLYAQYILQSLKR.  225
SEQ ID NO 112.pro MDVLVIMQEPVLPGCFLRAKAIGLMPMIDQGEKDDKIIAVCADDPEYRHYNDIKELPPHRLAEIRRFFEDYKKNENKEVAVNDFLPASAAYEAIKHSMTLYAEYVVENLRR.  217
SEQ ID NO 92.pro  MDVLVLMQEPVIPGSFLRARAIGLMPMIDQGEKDDKIIAVCADDPEYRHYNDISELSPHRLQEIKRFFEDYKKNENKEVAVDAFLPANTARDAIQYSMDLYAQYILQSLRQ.  214
```

Fig.2

Percent Identity

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35		
1		70.9	87.8	79.3	88.7	86.9	92.0	91.5	95.7	87.8	71.8	82.2	96.2	71.8	82.2	95.7	77.5	77.9	77.5	89.2	79.8	78.4	80.3	82.2	82.2	84.0	77.7	83.9	76.1	79.8	77.9	77.9	82.6	80.8	81.7	**1**	SEQ ID NO 30
2	35.2		71.0	71.9	69.1	69.6	70.2	71.6	73.4	69.6	92.2	72.0	71.9	92.2	72.5	73.4	71.7	78.0	71.7	72.3	71.4	79.0	72.1	71.6	72.1	70.8	75.7	73.7	72.6	72.4	69.5	67.9	69.9	72.4	72.9	**2**	SEQ ID NO 32
3	12.3	32.1		75.6	81.6	94.0	84.7	85.1	87.9	95.4	71.9	77.9	88.1	71.4	77.4	87.9	73.7	78.5	74.7	86.4	77.4	79.0	77.7	81.4	81.4	81.6	78.2	82.4	75.8	76.0	77.0	76.0	83.9	77.4	81.8	**3**	SEQ ID NO 34
4	20.2	30.3	24.4		75.6	75.6	77.7	78.6	81.6	75.1	71.9	90.3	80.5	72.4	92.2	81.6	82.0	78.0	80.2	78.4	85.3	78.0	83.7	77.2	77.2	78.3	80.7	78.5	74.9	82.9	75.1	80.2	76.0	85.3	78.0	**4**	SEQ ID NO 36
5	9.6	36.5	20.7	23.7		81.1	88.8	90.2	89.9	81.1	70.5	78.8	90.0	70.0	77.9	89.9	75.1	75.2	74.2	83.6	75.1	74.8	77.7	77.2	77.2	78.8	75.2	82.0	73.0	76.0	75.6	75.1	79.3	76.5	76.6	**5**	SEQ ID NO 38
6	14.0	34.2	5.8	25.1	21.3		84.2	84.7	87.9	97.2	70.0	77.0	88.6	69.6	77.0	87.9	73.7	77.6	74.7	84.5	76.0	78.0	76.7	81.4	81.4	80.6	78.2	82.9	75.3	76.0	76.1	76.0	82.0	76.0	81.3	**6**	SEQ ID NO 40
7	6.4	36.0	15.6	22.5	9.5	16.7		95.3	91.3	85.1	71.2	80.0	91.9	70.7	79.5	91.3	78.1	75.7	77.7	86.4	79.1	76.2	80.9	80.9	81.9	81.4	77.7	82.0	75.3	80.5	78.4	77.7	80.5	79.5	80.4	**7**	SEQ ID NO 42
8	6.9	33.7	15.0	21.2	7.9	16.1	4.3		90.8	84.7	72.6	81.4	91.4	72.1	80.9	90.8	77.7	76.6	77.2	86.9	79.1	77.6	80.5	81.4	82.3	80.9	77.2	84.4	75.8	80.0	78.4	77.7	81.4	80.5	79.9	**8**	SEQ ID NO 44
9	2.5	31.5	11.5	18.5	8.8	12.1	7.1	7.7		88.4	74.4	83.6	96.1	74.4	83.6	99.5	79.2	80.2	79.2	88.9	81.6	80.7	81.6	84.1	84.1	87.0	78.2	82.9	78.3	81.2	80.7	79.2	83.6	83.1	84.5	**9**	SEQ ID NO 46
10	12.9	34.2	4.3	25.1	21.3	2.4	15.6	16.1	11.0		70.0	77.0	89.5	69.6	77.0	88.4	73.3	77.6	74.2	85.4	76.5	78.0	77.2	80.5	80.5	81.1	77.7	82.0	74.9	77.0	76.5	75.6	82.5	76.5	81.3	**10**	SEQ ID NO 48
11	33.7	6.2	31.3	29.4	35.7	33.5	34.5	32.2	30.1	33.5		72.9	72.9	99.1	72.9	74.4	72.3	78.0	71.4	73.2	72.4	79.0	72.6	72.6	73.0	71.4	75.7	73.2	74.0	71.4	70.4	69.3	70.0	72.8	73.4	**11**	SEQ ID NO 50
12	17.0	30.3	21.1	8.3	20.5	23.1	19.3	17.5	16.6	22.4	29.4		83.3	72.9	95.9	83.6	80.7	76.6	80.7	81.2	83.4	77.1	83.7	78.6	78.6	79.8	80.7	80.0	74.9	81.1	75.1	81.2	76.6	84.3	78.0	**12**	SEQ ID NO 52
13	1.9	33.4	11.9	19.1	8.6	12.5	7.0	7.0	2.5	11.4	31.2	16.0		72.9	83.3	96.1	78.1	79.0	79.0	89.0	80.5	79.5	81.0	82.9	82.9	85.7	78.2	83.9	77.1	81.0	79.5	78.6	83.3	81.4	82.4	**13**	SEQ ID NO 54
14	33.7	6.2	32.1	29.4	35.7	34.2	35.2	33.0	30.1	34.2	0.5	29.4	31.9		72.9	74.4	72.3	78.0	71.4	73.2	72.4	79.0	72.6	72.1	72.6	71.4	75.7	73.2	74.0	71.9	70.4	69.3	70.0	72.8	73.4	**14**	SEQ ID NO 56
15	17.0	29.6	21.8	6.3	21.1	23.1	20.0	18.1	16.6	22.4	28.0	3.8	16.0	28.0		83.6	80.7	77.1	80.3	80.8	83.9	77.6	83.7	78.1	78.1	79.8	80.2	79.5	75.8	82.0	75.1	81.7	76.6	84.8	77.6	**15**	SEQ ID NO 58
16	2.5	31.5	11.5	18.5	8.8	12.1	7.1	7.7	0.0	11.0	30.1	16.6	2.5	30.1	16.6		79.2	80.2	79.2	88.9	81.6	80.7	81.6	84.1	84.1	87.0	78.2	82.9	78.3	81.2	80.7	79.2	83.6	83.1	84.5	**16**	SEQ ID NO 60
17	24.6	33.3	30.1	19.1	28.7	30.8	24.4	25.0	22.1	30.8	33.1	19.6	23.1	33.1	18.4	22.1		78.0	96.1	78.9	83.4	79.0	83.3	74.9	75.8	72.4	77.2	78.5	74.0	82.5	70.9	77.1	72.4	84.3	75.2	**17**	SEQ ID NO 62
18	24.6	23.6	24.6	20.7	26.6	25.3	26.6	25.3	21.9	25.3	25.5	22.0	22.9	25.5	22.0	21.9	23.2		79.0	78.4	78.0	97.7	77.6	78.0	79.0	79.0	80.7	78.0	78.0	75.7	73.7	72.4	79.0	79.0	79.9	**18**	SEQ ID NO 64
19	26.3	34.3	31.1	20.7	30.4	31.8	26.0	26.7	23.9	31.8	35.5	20.0	24.8	35.5	20.6	23.9	2.2	22.9		78.4	81.1	79.0	83.7	74.9	75.8	72.4	77.2	78.0	73.5	81.6	70.4	77.5	71.6	82.0	75.7	**19**	SEQ ID NO 66
20	9.7	31.0	13.4	20.2	16.9	15.7	12.9	11.8	9.4	14.6	28.9	16.4	10.3	29.6	17.0	9.4	23.3	22.9	24.4		79.3	78.9	82.6	81.2	82.2	83.1	78.7	83.9	77.0	80.8	77.9	77.9	83.6	80.8	82.2	**20**	SEQ ID NO 68
21	19.5	30.3	23.2	15.5	23.9	24.5	20.6	20.6	19.1	23.9	30.8	16.6	18.5	30.8	16.0	19.1	19.1	20.1	20.1	18.9		79.0	81.9	75.8	75.8	77.0	80.7	79.0	76.7	82.0	73.2	78.3	75.1	97.7	75.7	**21**	SEQ ID NO 70
22	24.0	22.3	24.0	20.7	27.3	24.6	26.0	24.0	21.2	24.6	24.2	21.3	22.3	24.2	21.3	21.2	21.9	1.9	22.9	22.3	18.8		78.0	78.0	79.0	79.0	80.7	79.5	79.0	76.6	73.7	72.9	79.0	79.9	79.9	**22**	SEQ ID NO 72
23	17.7	27.8	21.9	16.7	21.9	23.2	17.6	18.2	17.3	22.5	28.3	16.7	17.2	28.3	16.7	17.3	16.3	20.3	16.7	15.8	18.0	19.6		77.2	78.6	79.1	81.2	81.5	78.6	93.0	75.1	86.0	77.7	82.8	78.5	**23**	SEQ ID NO 74
24	18.4	31.0	19.2	24.4	25.5	19.2	21.9	21.3	15.9	20.4	31.7	23.1	18.0	31.7	23.7	15.9	28.4	21.6	29.4	18.9	26.6	21.0	21.9		97.7	86.5	79.7	83.4	77.7	76.3	81.7	75.3	86.5	76.3	93.9	**24**	SEQ ID NO 76
25	18.4	30.3	19.2	23.7	25.5	19.2	20.7	20.0	15.9	20.4	31.7	22.4	18.0	31.7	23.1	15.9	27.0	20.4	28.0	17.7	25.9	19.7	20.6	1.9		85.6	80.2	84.4	78.6	77.2	81.2	75.3	86.0	76.7	94.4	**25**	SEQ ID NO 78
26	15.8	35.2	17.8	25.2	23.3	20.2	19.2	19.8	13.3	18.4	35.8	23.2	15.5	35.8	23.2	13.3	32.8	24.4	33.8	17.0	25.2	24.4	24.2	13.3	14.4		81.7	82.4	76.7	77.0	82.6	74.3	83.6	77.9	86.0	**26**	SEQ ID NO 80
27	24.7	27.4	24.6	20.0	28.1	24.6	24.7	25.4	24.0	25.2	27.4	20.6	24.0	27.4	20.6	24.0	25.4	20.0	25.9	23.9	20.0	20.0	19.3	22.5	21.8	19.9		77.2	86.1	80.2	74.8	79.7	78.7	79.7	79.7	**27**	SEQ ID NO 82
28	16.5	30.4	17.6	21.8	18.9	17.0	18.4	15.3	17.1	18.3	31.2	19.9	16.0	31.2	20.5	17.1	22.4	23.6	24.1	16.5	21.2	21.6	17.9	17.0	15.2	18.9	25.2		79.0	81.0	80.0	79.0	84.4	80.0	83.4	**28**	SEQ ID NO 84
29	26.6	32.2	25.8	25.3	29.2	27.2	27.3	26.6	23.9	27.8	30.1	25.3	24.9	30.1	24.0	23.9	29.7	24.9	31.5	24.1	22.7	23.6	19.6	23.4	21.5	26.9	13.0	22.1		78.1	75.1	74.0	76.7	76.7	79.0	**29**	SEQ ID NO 86
30	18.9	29.0	24.4	18.4	23.7	24.4	18.7	20.0	18.5	23.7	29.6	20.1	17.8	29.6	18.9	18.5	17.9	22.8	18.9	17.6	19.0	21.5	4.3	23.9	22.5	27.2	20.6	19.2	21.5		74.2	86.2	75.6	82.9	77.6	**30**	SEQ ID NO 88
31	22.2	36.5	23.8	24.3	27.8	25.2	22.7	22.7	19.4	24.5	35.7	24.9	21.8	36.5	24.9	19.4	33.7	29.4	34.0	22.8	27.0	29.4	23.9	16.0	16.6	17.0	28.2	20.3	29.9	25.8		70.9	81.7	74.2	84.0	**31**	SEQ ID NO 90
32	21.5	35.4	25.1	20.7	25.1	24.4	21.9	21.9	21.7	25.1	35.2	21.2	21.1	35.2	20.6	21.7	24.0	28.8	25.0	20.8	23.2	28.1	13.3	27.1	26.4	31.1	21.3	21.8	26.0	13.7	30.6		72.0	77.9	74.8	**32**	SEQ ID NO 94
33	18.2	36.0	17.2	26.7	23.3	18.4	21.0	20.4	15.7	18.4	36.5	25.9	18.4	36.5	25.9	15.7	34.2	23.7	34.5	16.9	28.0	23.7	25.5	12.2	12.7	16.0	23.9	15.8	26.2	29.4	17.0	32.7		76.5	87.4	**33**	SEQ ID NO 96
34	18.3	28.9	22.5	15.5	22.5	23.9	20.6	19.3	17.8	23.2	29.4	15.4	17.2	29.4	14.8	17.8	17.9	18.8	18.9	17.6	1.9	17.6	16.8	25.2	24.5	24.0	21.3	19.9	22.7	17.8	25.6	23.8	26.7		76.2	**34**	SEQ ID NO 11
35	17.9	29.7	18.7	23.1	25.7	19.3	20.1	20.8	14.7	19.3	29.7	23.1	18.1	29.7	23.7	14.7	27.2	19.8	28.2	17.8	26.6	19.8	20.6	4.4	4.4	13.4	22.5	16.4	22.9	22.5	14.9	27.1	11.1	25.2		**35**	SEQ ID NO 92
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**		

WO 2011/008510
PCT/US2010/040281
CA 02764073 2011-11-30

FIG.3

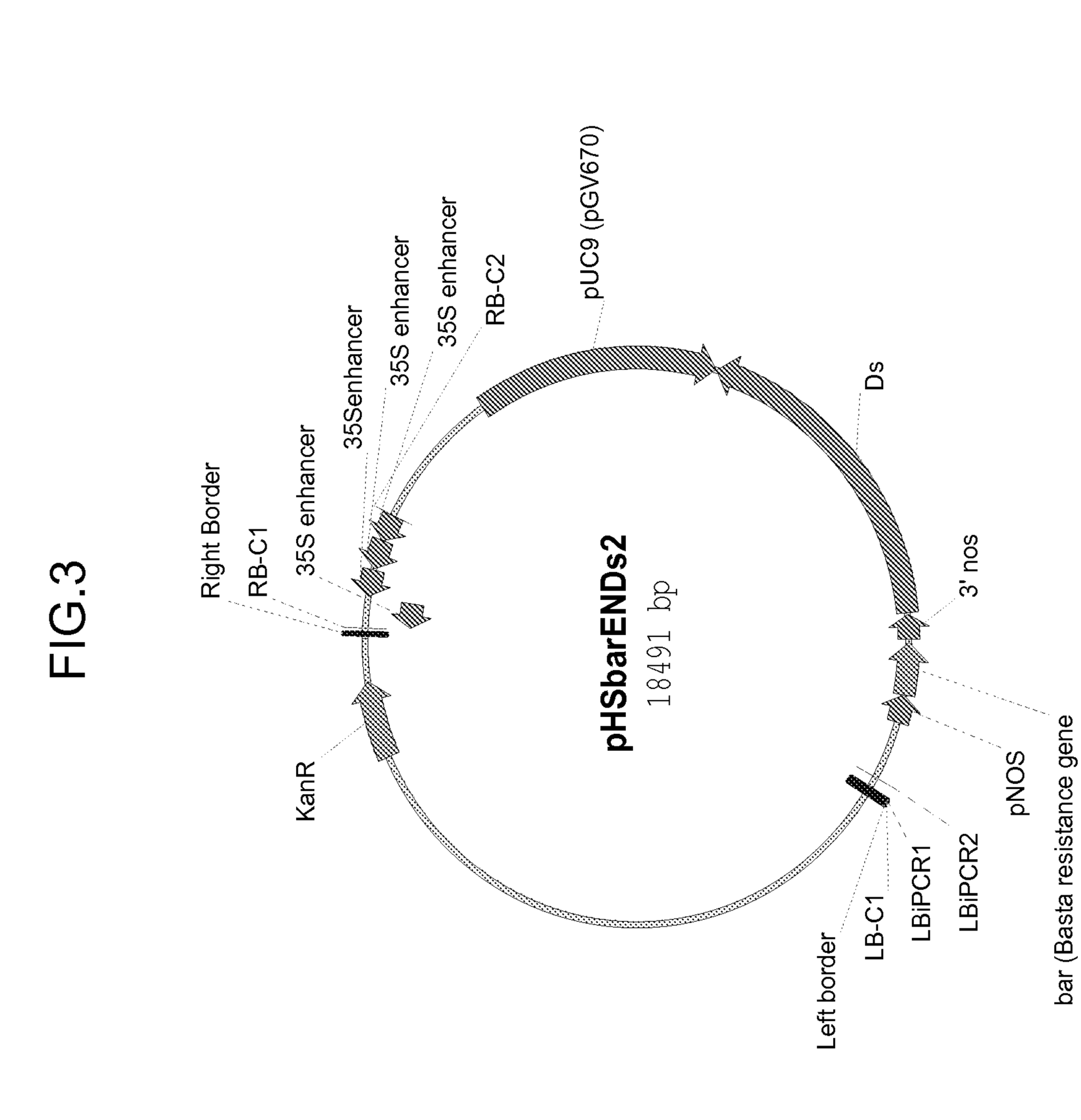
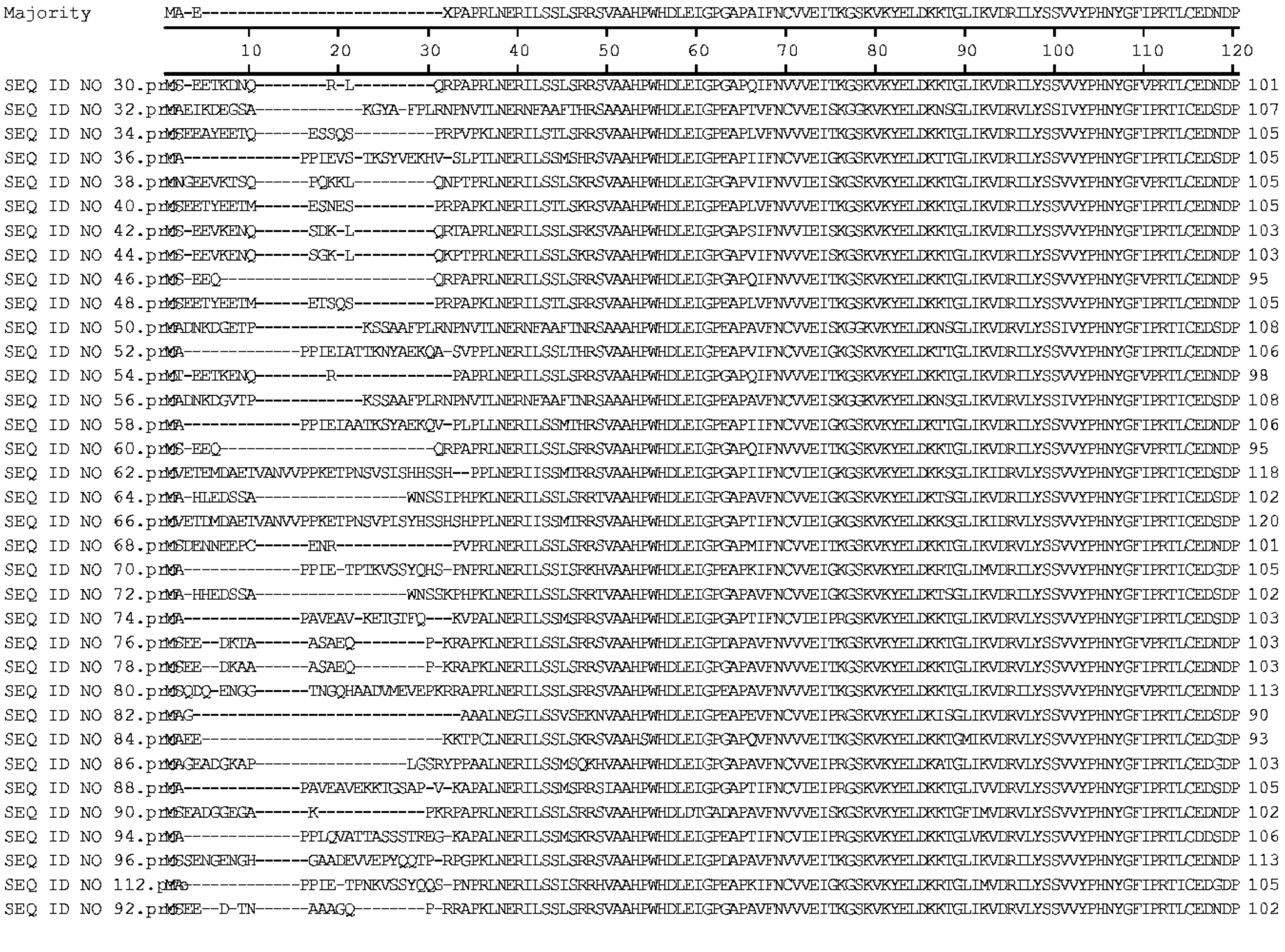

pHSbarENDs2
18491 bp
Right Border
RB-C1
35S enhancer
35Senhancer
35S enhancer
35S enhancer
RB-C2
pUC9 (pGV670)
Ds
3' nos
bar (Basta resistance gene
pNOS
LBiPCR2
LBiPCR1
LB-C1
Left border
KanR

FIG. 1A


Majority MA-E----------XPAPRLNERILSSLSRRSVAAHPWHDLEIGPGAPAIFNCVVEITKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFIPRTLCEDNDP
10 20 30 40 50 60 70 80 90 100 110 120
SEQ ID NO 30.pr MS-EETKDNQ-------R-L--------QRPAPRLNERILSSLSRRSVAAHPWHDLEIGPGAPQIFNVVVEITKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFVPRTLCEDNDP 101
SEQ ID NO 32.pr MAEIKDEGSA----------KGYA-FPLRNPNVTLNERNFAAFTHRSAAAHPWHDLEIGPEAPTVFNCVVEISKGGKVKYELDKNSGLIKVDRVLYSSIVYPHNYGFIPRTICEDSDP 107
SEQ ID NO 34.pr MSEEAYEETQ-----ESSQS--------PRPVPKLNERILSTLSRRSVAAHPWHDLEIGPEAPLVFNVVVEITKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFIPRTLCEDNDP 105
SEQ ID NO 36.pr MA----------PPIEVS-TKSYVEKHV-SLPTLNERILSSMSHRSVAAHPWHDLEIGPEAPIIFNCVVEIGKGSKVKYELDKTTGLIKVDRILYSSVVYPHNYGFIPRTLCEDSDP 105
SEQ ID NO 38.pr MNGEEVKTSQ-----PQKKL--------QNPTPRLNERILSSLSKRSVAAHPWHDLEIGPGAPVIFNVVIEISKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFVPRTLCEDNDP 105
SEQ ID NO 40.pr MSEETYEETM-----ESNES--------PRPAPKLNERILSTLSKRSVAAHPWHDLEIGPEAPLVFNVVVEITKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFIPRTLCEDNDP 105
SEQ ID NO 42.pr MS-EEVKENQ-----SDK-L--------QRTAPRLNERILSSLSRKSVAAHPWHDLEIGPGAPSIFNVVIEISKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFIPRTLCEDNDP 103
SEQ ID NO 44.pr MS-EEVKENQ-----SGK-L--------QKPTPRLNERILSSLSKRSVAAHPWHDLEIGPGAPVIFNVVVEISKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFIPRTLCEDNDP 103
SEQ ID NO 46.pr MS-EEQ----------QRPAPRLNERILSSLSRRSVAAHPWHDLEIGPGAPQIFNVVVEITKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFVPRTLCEDNDP 95
SEQ ID NO 48.pr MSEETYEETM-----ETSQS--------PRPAPKLNERILSTLSRRSVAAHPWHDLEIGPEAPLVFNVVVEITKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFIPRTLCEDNDP 105
SEQ ID NO 50.pr MADNKDGETP----------KSSAAFPLRNPNVTLNERNFAAFTNRSAAAHPWHDLEIGPEAPAVFNCVVEISKGGKVKYELDKNSGLIKVDRVLYSSIVYPHNYGFIPRTICEDSDP 108
SEQ ID NO 52.pr MA----------PPIEIATTKNYAEKQA-SVPPLNERILSSLTHRSVAAHPWHDLEIGPEAPVIFNCVVEIGKGSKVKYELDKTTGLIKVDRILYSSVVYPHNYGFIPRTLCEDNDP 106
SEQ ID NO 54.pr MT-EETKENQ-------R----------PAPRLNERILSSLSRRSVAAHPWHDLEIGPGAPQIFNVVVEITKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFVPRTLCEDNDP 98
SEQ ID NO 56.pr MADNKDGVTP----------KSSAAFPLRNPNVTLNERNFAAFTNRSAAAHPWHDLEIGPEAPAVFNCVVEISKGGKVKYELDKNSGLIKVDRVLYSSIVYPHNYGFIPRTICEDSDP 108
SEQ ID NO 58.pr MA----------PPIEIAATKSYAEKQV-PLPLLNERILSSMTHRSVAAHPWHDLEIGPEAPIIFNCVVEIGKGSKVKYELDKTTGLIKVDRILYSSVVYPHNYGFIPRTLCEDNDP 106
SEQ ID NO 60.pr MS-EEQ----------QRPAPRLNERILSSLSRRSVAAHPWHDLEIGPGAPQIFNVVVEITKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFVPRTLCEDNDP 95
SEQ ID NO 62.pr MVETEMDAETVANVVPPKETPNSVSISHHSSH--PPLNERIISSMIRRSVAAHPWHDLEIGPGAPIIFNCVIEIGKGSKVKYELDKKSGLIKIDRVLYSSVVYPHNYGFIPRTICEDSDP 118
SEQ ID NO 64.pr MA-HLEDSSA----------WNSSIPHPKLNERILSSLSRRTVAAHPWHDLEIGPGAPAVFNCVVEIGKGSKVKYELDKTSGLIKVDRILYSSVVYPHNYGFIPRTICEDSDP 102
SEQ ID NO 66.pr MVETDMDAETVANVVPPKETPNSVPISYHSSHSHPPLNERIISSMIRRSVAAHPWHDLEIGPGAPTIFNCVIEIGKGSKVKYELDKKSGLIKIDRVLYSSVVYPHNYGFIPRTICEDSDP 120
SEQ ID NO 68.pr MSDENNEEPC------ENR----------PVPRLNERILSSLSRRSVAAHPWHDLEIGPGAPMIFNCVVEITKGSKVKYELDKKTGLIKVDRVLYSSVVYPHNYGFIPRTLCEDNDP 101
SEQ ID NO 70.pr MA----------PPIE-TPTKVSSYQHS-PNPRLNERILSSISRKHVAAHPWHDLEIGPEAPKIFNCVVEIGKGSKVKYELDKRTGLIMVDRILYSSVVYPHNYGFIPRTICEDGDP 105
SEQ ID NO 72.pr MA-HHEDSSA----------WNSSKPHPKLNERILSSLSRRTVAAHPWHDLEIGPGAPAVFNCVVEIGKGSKVKYELDKTSGLIKVDRILYSSVVYPHNYGFIPRTICEDSDP 102
SEQ ID NO 74.pr MA----------PAVEAV-KETGTFQ---KVPALNERILSSMSRRSVAAHPWHDLEIGPGAPTIFNCVIEIPRGSKVKYELDKKTGLIKVDRVLYSSVVYPHNYGFIPRTLCEDSDP 103
SEQ ID NO 76.pr MSEE--DKTA------ASAEQ--------P-KRAPKLNERILSSLSRRSVAAHPWHDLEIGPDAPAVFNVVVEITKGSKVKYELDKKTGLIKVDRVLYSSVVYPHNYGFVPRTLCEDNDP 103
SEQ ID NO 78.pr MSEE--DKAA------ASAEQ--------P-KRAPKLNERILSSLSRRSVAAHPWHDLEIGPGAPAVFNVVVEITKGSKVKYELDKKTGLIKVDRVLYSSVVYPHNYGFIPRTLCEDNDP 103
SEQ ID NO 80.pr MSQDQ-ENGG------TNGQHAADVMEVEPKRRAPRLNERILSSLSRRSVAAHPWHDLEIGPEAPAVFNVVVEITKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFVPRTLCEDNDP 113
SEQ ID NO 82.pr MAG----------AAALNEGILSSVSEKNVAAHPWHDLEIGPEAPEVFNCVVEIPRGSKVKYELDKISGLIKVDRVLYSSVVYPHNYGFIPRTLCEDSDP 90
SEQ ID NO 84.pr MAEE----------KKTPCLNERILSSLSKRSVAAHSWHDLEIGPGAPQVFNVVVEITKGSKVKYELDKKTGMIKVDRVLYSSVVYPHNYGFIPRTLCEDGDP 93
SEQ ID NO 86.pr MAGEADGKAP----------LGSRYPPAALNERILSSMSQKHVAAHPWHDLEIGPGAPAVFNCVVEIPRGSKVKYELDKATGLIKVDRVLYSSVVYPHNYGFIPRTLCEDGDP 103
SEQ ID NO 88.pr MA----------PAVEAVEKKTGSAP-V-KAPALNERILSSMSRRSIAAHPWHDLEIGPGAPTIFNCVIEIPRGSKVKYELDKKTGLIVVDRVLYSSVVYPHNYGFIPRTLCEDSDP 105
SEQ ID NO 90.pr MSEADGGEGA------K----------PKRPAPRLNERILSSLSRRSVAAHPWHDLDTGADAPAVFNVVVEISKGSKVKYELDKKTGFIMVDRVLYSSVVYPHNYGFIPRTLCEDNDP 102
SEQ ID NO 94.pr MA----------PPLQVATTASSSTREG-KAPALNERILSSMSKRSVAAHPWHDLEIGPEAPTIFNCVIEIPRGSKVKYELDKKTGLVKVDRVLYSSVVYPHNYGFIPRTLCDDSDP 106
SEQ ID NO 96.pr MSSENGENGH------GAADEVVEPYQQTP-RPGPKLNERILSSLSRRSVAAHPWHDLEIGPDAPAVFNVVVEITKGSKVKYELDKKTGLIKVDRILYSSVVYPHNYGFIPRTLCEDNDP 113
SEQ ID NO 112.p MA----------PPIE-TPNKVSSYQQS-PNPRLNERILSSISRRHVAAHPWHDLEIGPEAPKIFNCVVEIGKGSKVKYELDKRTGLIMVDRILYSSVVYPHNYGFIPRTICEDGDP 105
SEQ ID NO 92.pr MSEE--D-TN------AAAGQ--------P-RRAPKLNERILSSLSRRSVAAHPWHDLEIGPGAPAVFNVVVEITKGSKVKYELDKKTGLIKVDRVLYSSVVYPHNYGFIPRTLCEDNDP 102