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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 May 2011 (05.05.2011)

PCT

(10) International Publication Number
WO 2011/053898 A2

(51) International Patent Classification:
C12N 15/82 (2006.01)

(21) International Application Number:
PCT/US2010/054932

(22) International Filing Date:
1 November 2010 (01.11.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/256,323 30 October 2009 (30.10.2009) 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, Wilmington, Delaware 19809 (US). **STECCA, Kevin, L.** [US/US]; 34 Phoebe Farms Lane, New Castle, Delaware 19720 (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, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

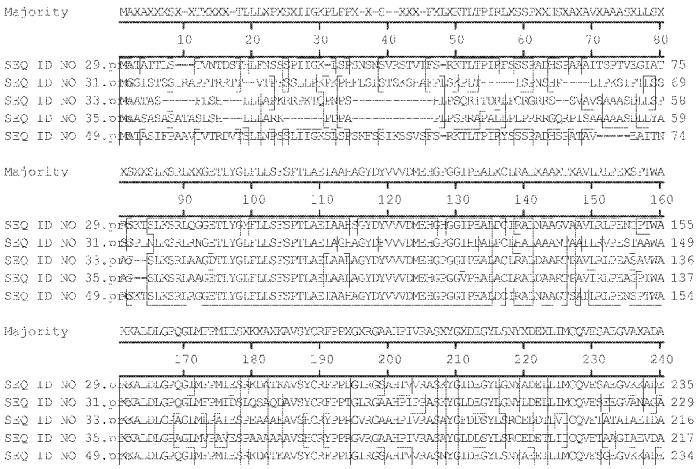
— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

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

WO 2011/053898 A2

(54) **Title**: PLANTS AND SEEDS WITH ALTERED STORAGE COMPOUND LEVELS, RELATED CONSTRUCTS AND METHODS INVOLVING GENES ENCODING PROTEINS WITH SIMILARITY TO BACTERIAL 2,4-DIHYDROXY-HEPT-2-ENE-1,7-DIOIC ACID CLASS II-LIKE ALDOLASE PROTEINS

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 proteins with similarity to bacterial 2,4-dihydroxy-hept-2-ene-1,7-dioic acid class II-like aldolase proteins in plants and seeds and the use of such fragments to modulate expression of a gene encoding proteins with similarity to bacterial 2,4-dihydroxy-hept-2-ene-1,7-dioic acid class II-like aldolase proteins in a transformed host cell.

TITLE

PLANTS AND SEEDS WITH ALTERED STORAGE COMPOUND LEVELS, RELATED CONSTRUCTS AND METHODS INVOLVING GENES ENCODING PROTEINS WITH SIMILARITY TO BACTERIAL 2,4-DIHYDROXY-HEPT-2-ENE-1,7-DIOIC ACID CLASS II-LIKE ALDOLASE PROTEINS

This application claims priority benefit of provisional application no. 61/256,323 filed October 30, 2009, the entire contents of which are hereby incorporated by reference.

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 proteins with similarity to bacterial 2,4-dihydroxy-hept-2-ene-1,7-diolic acid class II-like aldolase proteins in plants and seeds and the use of such fragments to modulate expression of a gene encoding plastidic ClassII aldolase-like 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.

Aldolases represent a diverse class of enzymes that differ in their catalytic mechanism and carbonyl donor preference (Wang et al. Biochemistry:44, 9447-9455 (2005)). There are Class I and Class II aldolases. Class II aldolases can be further devided into those that have a preference for dihydroxyacetone phosphate (DHAP) and those that prefer pyruvate as the carbonyl donor. The former represent the best characterized subgroup of Class II aldolases and includes for example fructose -1,6-bisphosphate aldolase, which catalyzes the cleavage of fructose 1,6-bisphosphate into two interconvertable three-carbon fragments: D-glyceraldehyde 3-phosphate and dihydroxyacetone phosphate ,which is the third committed step in glycolysis.

Class II pyruvate-specific aldolases include for example HpaI, a bacterial class II aldolase that catalyzes the reversible cleavage of 2,4-dihydroxy-hept-2-ene-1,7-dioic acid to pyruvate and succinic semialdehyde.

No studies on plant enzymes with similarity to bacterial 2,4-dihydroxy-hept-2-ene-1,7-dioic class II-like aldolase have been conducted and further investigation of the role of this subgroup of proteins in the regulation of storage compounds is therefore merited.

Diacylglycerol acyltransferase ("DGAT") is an integral membrane protein that catalyzes the final enzymatic step in the production of triacylglycerols in plants, fungi and mammals. This enzyme is responsible for transferring an acyl group from acyl-

coenzyme-A to the sn-3 position of 1,2-diacylglycerol ("DAG") to form triacylglycerol ("TAG"). DGAT is associated with membrane and lipid body fractions in plants and fungi, particularly, in oilseeds where it contributes to the storage of carbon used as energy reserves. TAG is believed to be an important chemical for storage of energy in cells. DGAT is known to regulate TAG structure an direct TAG synthesis. Furthermore, it is known that the DGAT reaction is specific for oil synthesis.

TAG is the primary component of vegetable oil in plants, It is used by the seed as a stored form of energy to be used during seed germination.

Two different families of DGAT proteins have been identified. The first family of DGAT proteins ("DGAT1") is related to the acyl-coenzyme A:cholesterol acyltransferase ("ACAT") and has been described in U.S. Patent Nos. 6,100,077 and 6,344,548. A second family of DGAT proteins ("DGAT2") is unrelated to the DGAT1 family and is described in PCT Patent Publication WO 2004/011671 published February 5, 2004. Other references to DGAT genes and their use in plants include PCT Publication Nos.WO2004/011,671, WO1998/055,631, and WO2000/001,713, and US Patent Publication No. 20030115632.

Applicants' Assignee's copending published patent application US 2006-0094088 describes genes for DGATs of plants and fungi and their use is in modifying levels of polyunsaturated fatty acids ("PUFAs") in edible oils.

Applicants' Assignee's published PCT application WO 2005/003322 describes the cloning of phosphatidylcholine diacylglycerol acyltransferase and DGAT2 for altering PUFA and oil content in oleaginous yeast.

Applicants' Assignee's copending published US application no. 12/470509 describes DGAT genes from Yarrpowia lipolytica combined with plastidic phosphoglucomutase down regulation for increased seed storage lipid production and altered fatty acid profiles in oilseed plants.

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 85% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 29, 31, 33, 35, 49, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120,

121, 122, 123, and 147and wherein seeds from said transgenic plant have an altered oil, protein, starch and/or soluble carbohydrate content and/or altered seed weight, 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 85% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 29, 31, 33, 35, 49, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, and 147 and wherein said transgenic seed has an altered oil, protein, starch and/or soluble carbohydrate content and/or altered seed weight 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 85% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 29, 31, 33, 35, 49, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, and 147 and wherein said transgenic seed has an increased protein 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 85% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 29, 31, 33, 35, 49, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, and 147or (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 85% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 29, 31,

33, 35, 49, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, and 147 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 85% 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 plastidic HpaIL aldolase, and wherein said plant has an altered oil, protein, starch and/or soluble carbohydrate content and/or altered seed weight 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: 28, 30, 32, 34, 48, 124, 125, 126,127, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, or 146; or (b) the full-length complement of (a): wherein (a) or (b) is of sufficient length to inhibit expression of endogenous plastidic HpaIL aldolase 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: 28, 30, 32, 34, 48, 124, 125, 126,127, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, or 146; or (b) the full-length complement of (a): wherein (a) or (b) is of sufficient length to inhibit expression of endogenous plastidic HpaIL aldolase 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 85% sequence identity,

based on the Clustal V method of alignment, when compared to SEQ ID NO: 29, 31, 33, 35, 49, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, and 123; 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 and/or altered seed weight, 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 85% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 29, 31, 33, 35, 49, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, and 123; 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 protein 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: 28, 30, 32, 34, 48, 124, 125, 126,127, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, or 146; or (ii) the full-length complement of (i); wherein (i) or (ii) is of sufficient length to inhibit expression of endogenous plastidic HpaIL aldolase 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 and/or altered seed weight, 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: 28, 30, 32, 34, 48, 124, 125, 126,127, 133, 134, 135, 136,

137, 138, 139, 140, 141, 142, 143, 144, 145, or 146; or (ii) the full-length complement of (i); wherein (i) or (ii) is of sufficient length to inhibit expression of endogenous plastidic HpaIL aldolase 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.

In a tenth embodiment this invention concerns transgenic plants comprising at least one DGAT sequences and a construct downregulating plastidic HpaI or HpaI-like activity, wherein the DGAT sequence and the plastidic HpaI or HpaI-like construct can be in the same recombinant construct or in separate recombinant constructs, and wherein seed obtained from said transgenic plant has an increased oil content when compared to the oil content of seed obtained from a control plant not comprising said construct or when compared to transgenic seed obtained from a transgenic plant comprising either said DGAT sequences alone or said construct downregulating HpaIL activity alone.

In an eleventh embodiment this invention concerns transgenic seed obtained from the transgenic plant comprising at least one DGAT sequence and a construct downregulating HpaIL activity, wherein the DGAT sequence and the plastidic HpaI or HpaI-like construct can be in the same recombinant construct or in separate recombinant constructs and wherein the oil content of said transgenic seed is increased when compared to the oil content of control seed not comprising said construct or null segregant or transgenic seed comprising either said DGAT sequences alone or said construct downregulating HpaIL activity alone.

In a twelfth embodiment this invention concerns a method for increasing the oil content of a seed comprising: (a) transforming at least one cell with at least one recombinant construct having at least one DGAT sequences and a construct downregulating plastidic HpaIL activity wherein the DGAT sequence and HpaIL construct can be in the same recombinant construct or in separate recombinant constructs; (b) selecting the transformed soybean cell(s) of step (a) having an increased oil acid content when compared to the oil content of a control cell not comprising said construct or when compared to a null segregant seed or when

compared to transgenic seed obtained from a transgenic plant comprising either said DGAT sequences alone or said construct downregulating HpaIL activity alone.

In a thirteen's embodiment this invention concerns an isolated polynucleotide comprising: (a) a nucleotide sequence encoding a polypeptide with HpaIL aldolase activity, wherein, based on the Clustal V method of alignment with pairwise alignment default parameters of KTUPLE=1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5, the polypeptide has an amino acid sequence of at least 75% sequence identity when compared to SEQ ID NO:120, 121, 122 or 123, or (b)the full complement of the nucleotide sequence of (a).

In a fourteens embodiment this invention concerns an isolated polynucleotide comprising: (a) a nucleotide sequence encoding a polypeptide with HpaIL aldolase activity, wherein, based on the Clustal V method of alignment with pairwise alignment default parameters of KTUPLE=1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5, wherein the amino acid sequence of the polypeptide comprises SEQ ID NO: 120, 121, 122 or 123.

In a fifteens embodiment this invention concerns an isolated polynucleotide comprising: (a) a nucleotide sequence encoding a polypeptide with HpaIL aldolase activity, wherein, based on the Clustal V method of alignment with pairwise alignment default parameters of KTUPLE=1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5, wherein the nucleotide sequence comprises SEQ ID NO:124, 125, 126, or 127.

In a sixteen's embodiment this invention concerns an isolated polynucleotide encoding a polypeptide, wherein said polynucleotide is capable of altering the endogenous expression of plastidic HpaIL aldolase activity and wherein said polypeptide comprises a chloroplast transit peptide and at least one motif selected from the group consisting of: SEQ ID NO: 128, 129, 130, 131, or 132.

In a seventeenth embodiment this invention concerns an isolated polynucleotide encoding a plant HpaIL aldolase polypeptide, wherein said polynucleotide is capable of altering the endogenous expression of plastidic Hpal-like activity and wherein said polypeptide has a Km (acetaldehyde) at least 1.7 fold lower than the Km (acetaldehyde) of bacterial HpaIL aldolase activity and a Vmax of at least 15 fold lower than the bacterial HpaIL aldolase activity.

In an eighteenth embodiment this invention concerns a method of altering

i.e. increasing or decreasing oil, protein, starch and/or soluble carbohydrate content and or altering seed weight, comprising: a) transforming a plant with the recombinant DNA construct of claim 34; b) growing the transformed plant under conditions suitable for the expression of the recombinant DNA construct; and c) selecting those plant having altered i.e. increased or decreased oil, protein, starch and/or soluble carbohydrate content and/or altered seed weight when compared to a control plant not comprising said recombinant DNA construct.

In a nineteenth embodiment the present invention concerns a method to isolate nucleic acid fragments encoding plastidic HpaI-like polypeptides, comprising: a) comparing SEQ ID NOs: 128, 129, 130, 131, or 132 with other polypeptide sequences encoding a plastidic HpaI-like polypeptides; b) identifying the conserved sequences obtained in step (a); c) making region-specific nucleotide probe(s) or oligomer(s) based on the conserved sequences identified in step (b); and (d) using the nucleotide probe(s) or oligomer(s) of step (c) to isolate HpaI-like sequences; e) selecting those sequences comprising a chloroplast transit peptide.

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 plants or seed comprising the recombinant DNA construct of the present invention are useful to alter i.e. increase or decrease oil, protein, starch and/or soluble carbohydrate content and/or altered seed weight when compared to a control plant not comprising the recombinant DNA construct(s) of the present invention.

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 plastidic HpaIL aldolase activity, wherein the method comprises transforming a monocot or dicot plant with any of the nucleic acid fragments of the present invention.

Progeny plants derived from the transgenic plant, wherein said progeny plant comprises in its genome the recombinant DNA construct and seed obtained from said progeny plant exhibit an altered i.e. increased or decreased oil, protein, starch and/or soluble carbohydrate content and/or altered seed weight when compared to a control plant not comprising said recombinant DNA construct are also included in the present invention.

Furthermore the present invention includes a vector comprising any of the isolated polynucleotides of the present invention. Also included are methods for transforming a cell comprising transforming a cell with any of the isolated polynucleotides of the present invention. The cell transformed by this method is also included. Advantageously, the cell is eukaryotic, e.g., a yeast, insect or plant cell, or prokaryotic, e.g., a bacterium.

In another embodiment, the present invention includes a method for producing a transgenic plant comprising transforming a plant cell with any of the isolated polynucleotides or recombinant DNA constructs of the present invention and regenerating a transgenic plant from the transformed plant cell. The invention is also directed to the transgenic plant produced by this method, and transgenic seed obtained from this transgenic plant.

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.

FIGs. 1A-1B shows an alignment of the amino acid sequences of plastidic HpaIL aldolases encoded by the nucleotide sequences derived from the following: *Arabidopsis thaliana* (SEQ ID NO: 49); canola (SEQ ID NO:29); soybean (SEQ ID NO:31); corn (SEQ ID NO:33), and rice (SEQ ID NO:35). 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.

FIG.4A-C show an alignment of the amino acid sequences of plastidic HpaIL aldolases encoded by the nucleotide sequences derived from the following: *Arabidopsis lyrata* (SEQ ID NO: 107); Theobroma cacao (SEQ ID NO:108); Ricinus communis (SEQ ID NO:109); Solanum lycopersicum (SEQ ID NO:110), Vitis vinifera (SEQ ID NO:111), Carica papaya (SEQ ID NO:112), Citrus clementina (SEQ ID NOs:113 and 114), Oryza brachyata (SEQ ID NO:115), Brachypodium distayon (SEQ ID NO:116), Sorghum (SEQ ID NO:117), Paspalum notatum (SEQ ID NO:118), Eragrostis nindensis (SEQ ID NO:119), Tulipa gesneriana (SEQ ID NOs:120 and 121), Brassica napus (SEQ ID NO:29), Glycine max (SEQ ID NO:31), Zea Mays (SEQ ID NO:33), Oryza sativa (SEQ ID NO:35), and Arabidopsis thaliana (SEQ ID NO:49). Amino acids conserved among all sequences are indicated with an asterix above the conserved residues. Conservative amino acids substitutions are indicated by a plus sign (+) above the conserved residues. Dashes are used by the program to maximize alignment of the sequences. Conserved sequence motifs I, II, III, IV and V are underlined. The active site residue "R83" is indicated by a triangle under the alignment.

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 lo22048,

SEQ ID NO:11 corresponds to the forward primer HpaILORF FWD.

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

SEQ ID NO:13 corresponds to the nucleotide sequence of vector pENTR-HpaIL.

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

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 HpaIL UTR FWD forward primer.

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

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

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

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

Identification of plant genes with similarity to At4g10750 (HpaIL aldolases)

Protein (Plant Source)	Clone Designation	Status	SEQ ID NO: (Nucleotide)	(Amino Acid)
HpaIL aldolase *(Brassica napus)*	TC25873	CGS	28	29
HpaIL aldolase (*Glycine max)*	Glyma09g21760	CGS	30	31
HpaIL aldolase *(Zea mays)*	PCO651314	CGS	32	33
HpaIL aldolase *(Oryza sativa)*	Os09g36030	CGS	34	35
HpaIL aldolase *(Zea mays)*	cfp2npk070b11.fis1	CGS	146	147

SEQ ID NO:36 is the linker sequence described in Example 14.

SEQ ID NO:37 is the nucleic acid sequence of vector pKS133 described in Example 15.

SEQ ID NO:38 corresponds to synthetic complementary region of pKS106 and pKS124.

SEQ ID NO:39 corresponds to a synthetic complementary region of pKS133.

SEQ ID NO:40 corresponds to a synthetic PCR primer.

SEQ ID NO:41 corresponds to a synthetic PCR primer.

SEQ ID NO:42 corresponds to a synthetic PCR primer (SA64).

SEQ ID NO:43 corresponds to a synthetic PCR primer (SA65).

SEQ ID NO:44 corresponds to a synthetic PCR primer (SA66).

SEQ ID NO:45 is the nucleic acid sequence of vector pKS423.

SEQ ID NO:46 corresponds to the nucleic acid sequence of plasmid pKS120.

SEQ ID NO:47 corresponds to the nucleic acid sequence of At4g10750.

SEQ ID NO:48 corresponds to the ORF of SEQ ID NO:47.

SEQ ID NO:49 corresponds to the amino acid sequence encoded by SEQ ID NO:48.

SEQ ID NO:50 corresponds to the nucleotide sequence of pENTR-At4g10750.

SEQ ID NO:51 corresponds to the nucleotide sequence of pKR1478-At4g10750.

SEQ ID NO:52 corresponds to the nucleotide sequence of pKR1478-At4g10750-ORF.

SEQ ID NO:53 corresponds to the amino acid sequence encoded by the ORF in SEQ ID NO:52.

SEQ ID NO:54 corresponds to the nucleotide sequence of pKR1482-At4g10750.

SEQ ID NO:55 corresponds to the nucleotide sequence of KS387.

SEQ ID NO:56 corresponds to the nucleotide sequence artificial microRNA (amiRNA) aldo A.

SEQ ID NO:57 corresponds to the nucleotide sequence amiRNA aldo B.

SEQ ID NO:58 corresponds to the nucleotide sequence amiRNA aldo A star.

SEQ ID NO:59 corresponds to the nucleotide sequence amiRNA aldo B star.

SEQ ID NO:60 corresponds to the nucleotide sequence of microRNA 159 precursor.

SEQ ID NO:61 corresponds to the nucleotide sequence of in-fusion ready microRNA 159.

SEQ ID NO:62 corresponds to the nucleotide sequence of in-fusion ready microRNA 159-KS126 plasmid.

SEQ ID NO:63 corresponds to the nucleotide sequence of the gmir159ALDO A1 primer.

SEQ ID NO:64 corresponds to the nucleotide sequence of the gmir159ALDO A2 primer.

SEQ ID NO:65 corresponds to the nucleotide sequence of the 159-ALDO A DNA microRNA precursor.

SEQ ID NO:66 corresponds to the nucleotide sequence of the gmir159ALDO B1 primer.

SEQ ID NO:67 corresponds to the nucleotide sequence of the gmir159ALDO B2 primer.

SEQ ID NO:68 corresponds to the nucleotide sequence of the 159-ALDO B DNA microRNA precursor.

SEQ ID NO:69 corresponds to the nucleotide sequence of 159 ALDO A-KS126.

SEQ ID NO:70 corresponds to the nucleotide sequence of 159 ALDO B-KS126

SEQ ID NO:71 corresponds to the nucleotide sequence of the AthHpaIL fwd

primer.

SEQ ID NO:72 corresponds to the nucleotide sequence of the AthHpaIL rev primer.

SEQ ID NO:73 corresponds to the nucleotide sequence of pGEM®-T At4g10750.

SEQ ID NO:74 corresponds to the nucleotide sequence of pET28a At4g10750

SEQ ID NO:75 corresponds to the amino acid sequence of At4g10750 His TAG AA.

SEQ ID NO:76 corresponds to the nucleotide sequence of the Soy HpaIL fwd primer.

SEQ ID NO:77 corresponds to the nucleotide sequence of the Soy HpaIL rev primer.

SEQ ID NO:78 corresponds to the nucleotide sequence pGEM®-T Glyma09g21760.

SEQ ID NO:79 corresponds to the nucleotide sequence pET29a Glyma09g21760.

SEQ ID NO:80 corresponds to the amino acid sequence of Glyma09g21760 His TAG.

SEQ ID NO:81 corresponds to the nucleotide sequence of the Rice HpaIL fwd primer.

SEQ ID NO:82 corresponds to the nucleotide sequence of the Rice HpaIL rev primer.

SEQ ID NO:83 corresponds to the nucleotide sequence of pGEM®-T Os09g36030.

SEQ ID NO:84 corresponds to the nucleotide sequence of pET28a Os09g36030.

SEQ ID NO:85 corresponds to amino acid sequence of Os09g36030 His TAG.

SEQ ID NO:86 corresponds to the nucleotide sequence of the PP FWD primer.

SEQ ID NO:87 corresponds to the nucleotide sequence of the PP REV primer.

SEQ ID NO:88 corresponds to the nucleotide sequence of pCR blunt HpaI PP.

SEQ ID NO:89 corresponds to the nucleotide sequence HpaI PP.

SEQ ID NO:90 corresponds to the amino acid sequence of HpaI PP.

SEQ ID NO:91 corresponds to the nucleotide sequence of HpaI PP fwd primer.

SEQ ID NO:92 corresponds to nucleotide sequence of HpaI PP rev primer.

SEQ ID NO:93 corresponds to the nucleotide sequence of pGEM®-T HpaI PP.

SEQ ID NO:94 corresponds to the nucleotide sequence of pET29a HpaI PP.

SEQ ID NO:95 corresponds to the amino acid sequence of HpaI PP His TAG.

SEQ ID NO:96 corresponds to the nucleotide sequence of the AthHpaIL G83 rev primer.

SEQ ID NO:97 corresponds to the nucleotide sequence of the AthHpaIL G83 fwd primer.

SEQ ID NO:98 corresponds to the nucleotide sequence of pGEM®-T At4g10750-G83

SEQ ID NO:99 corresponds to the nucleotide sequence of pET28a At4g10750-G83.

SEQ ID NO:100 corresponds to the amino acid sequence of At4g10750-G83 His TAG.

SEQ ID NO:101 corresponds to the nucleotide sequence of the FUSION REV primer.

SEQ ID NO:102 corresponds to the nucleotide sequence of FUSION FWD primer.

SEQ ID NO:103 corresponds to the nucleotide sequence of pET29a 3primer.

SEQ ID NO:104 corresponds to the nucleotide sequence of pCR8GW- plastid HpaI PP.

SEQ ID NO:105 corresponds to the nucleotide sequence of pKR1478 - plastid HpaI PP.

SEQ ID NO:106 corresponds to the amino acid sequence of pKR1478 - plastid HpaI PP AA.

SEQ ID NO:107 corresponds to the amino acid sequence of the plastidic HpAIL from *Arabidopsis lyrata* (NCBI GI NO: 297809303).

SEQ ID NO:108 corresponds to the amino acid sequence of the plastidic HpAIL from *Theobroma cacao* (NCBI GI NO: 212319639).

SEQ ID NO:109 corresponds to the amino acid sequence of the plastidic HpaIL aldolase from *Ricinus communis* (NCBI GI NO: 255587508).

SEQ ID NO:110 corresponds to the amino acid sequence of the plastidic HpaIL aldolase from *Solanum lycopersicum* (NCBI GI NO: 47105574).

SEQ ID NO:111 corresponds to the amino acid sequence of the plastidic HpaIL aldolase from *Vitis vinifera* (NCBI GI NO: 225426623 and proprietary clone vpl1c.pk008.o13).

SEQ ID NO:112 corresponds to the amino acid sequence of the plastidic HpaIL aldolase from *Carica papaya* (C_papaya Tu SC 175.15).

SEQ ID NO:113 corresponds to the amino acid sequence of the plastidic HpaIL aldolase from *Citrus clementina* (NCBI GI NO: 110855269).

SEQ ID NO:114 corresponds to the amino acid sequence of the plastidic HpAIL

from *Citrus clementina* (NCBI GI NO: 110843679).

SEQ ID NO:115 corresponds to the amino acid sequence of the plastidic HpAIL from *Oryza brachyata* (NCBI GI NO: 110430657).

SEQ ID NO:116 corresponds to the amino acid sequence of the plastidic HpAIL from *Brachypodium distachyon* (Bradi4g35820).

SEQ ID NO:117 corresponds to the amino acid sequence of the plastidic HpAIL from *Sorghum* (Sb02g030560).

SEQ ID NO:118 corresponds to the amino acid sequence of the plastidic HpAIL from *Paspalum notatum* (Bahia-force joined).

SEQ ID NO:119 corresponds to the amino acid sequence of the plastidic HpAIL from *Eragrostis nindensis* (resurrection grass -force joined).

SEQ ID NO:120 corresponds to the amino acid sequence of the plastidic HpAIL from *Tulipa gesneriana* (proprietary clone etp1c.pk001.g3:fis).

SEQ ID NO:121 corresponds to the amino acid sequence of the plastidic HpAIL from *Tulipa gesneriana* (proprietary clone etp1c.pk003.b22:fis).

SEQ ID NO:122 corresponds to the amino acid sequence of the plastidic HpAIL from *Asclepias syriaca* (proprietary clone mas1c.pk012.d9.f).

SEQ ID NO:123 corresponds to the amino acid sequence of the plastidic HpAIL from *Momordica charantia* (proprietary clone fds1n.pk007.i18).

SEQ ID NO:124 corresponds to the nucleic acid sequence of the plastid HpAIL from *Tulipa gesneriana* (proprietary clone etp1c.pk001.g3:fis) encoding the amino acid sequence set forth in SEQ ID NO:120.

SEQ ID NO:125 corresponds to the nucleic acid sequence of the plastid HpAIL from *Tulipa gesneriana* (proprietary clone etp1c.pk003.b22:fis) encoding the amino acid sequence set forth in SEQ ID NO:121.

SEQ ID NO:126 corresponds to the nucleic acid sequence of the plastid HpAIL from *Asclepias syriaca* (proprietary clone mas1c.pk012.d9.f) encoding the amino acid sequence set forth in SEQ ID NO:122.

SEQ ID NO:127 corresponds to the nucleic acid sequence of the plastid HpAIL from *Momordica charantia* (proprietary clone fds1n.pk007.i18) encoding the amino acid sequence set forth in SEQ ID NO:123.

SEQ ID NO:128 is a conserved sequence motif useful in identifying genes belonging to the HpaIL family of genes.

SEQ ID NO:129 is a conserved sequence motif useful in identifying genes belonging to the HpaIL family of genes.

SEQ ID NO:130 is a conserved sequence motif useful in identifying genes belonging to the HpaIL family of genes.

SEQ ID NO:131 is a conserved sequence motif useful in identifying genes belonging to the HpaIL family of genes.

SEQ ID NO:132 is a conserved sequence motif useful in identifying genes belonging to the HpaIL family of genes.

SEQ ID NO:133 corresponds to the nucleic acid sequence of the plastidic HpAIL from *Arabidopsis lyrata* encoding SEQ ID NO:107.

SEQ ID NO:134 corresponds to the nucleic acid sequence of the plastidic HpAIL from *Theobroma cacao* encoding SEQ ID NO:108.

SEQ ID NO:135 corresponds to the nucleic acid sequence of the plastidic HpAIL from *Ricinus communis* encoding SEQ ID NO:109.

SEQ ID NO:136 corresponds to the nucleic acid sequence of the plastidic HpAIL from *Solanum lycopersicum* encoding SEQ ID NO:110.

SEQ ID NO:137 corresponds to the nucleic acid sequence of the plastidic HpAIL from *Vitis vinifera* encoding SEQ ID NO:111.

SEQ ID NO:138 corresponds to the nucleic acid sequence of the plastidic HpAIL from *Carica papaya* (C_papaya Tu SC 175.15) encoding SEQ ID NO:112.

SEQ ID NO:139 corresponds to the nucleic acid sequence of the plastidic HpAIL from *Citrus clementina* encoding SEQ ID NO:113.

SEQ ID NO:140 corresponds to the nucleic acid sequence of the plastidic HpAIL from *Citrus clementina* encoding SEQ ID NO:114.

SEQ ID NO:141 corresponds to the nucleic acid sequence of the plastidic HpAIL from *Oryza brachyata* encoding SEQ ID NO:115.

SEQ ID NO:142 corresponds to the nucleic acid sequence of the plastidic HpAIL from *Brachypodium distachyon* (Bradi4g35820) encoding SEQ ID NO:116.

SEQ ID NO:143 corresponds to the nucleic acid sequence of the plastidic HpAIL from *Sorghum* (Sb02g030560) encoding SEQ ID NO:117.

SEQ ID NO:144 corresponds to the nucleic acid sequence of the plastidic HpAIL from *Paspalum notatum* (Bahia-force joined) encoding SEQ ID NO:118.

SEQ ID NO:145 corresponds to the nucleic acid sequence of the plastidic HpAIL from *Eragrostis nindensis* (resurrection grass -force joined) encoding SEQ ID NO:119.

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.

“Diacylglycerol acyltransferase” is abbreviated DAG AT or DGAT.

“Diacylglycerol” is abbreviated DAG.

The term “HpaI-like aldolase”, “HpaIL”, “HpaIL aldolase” and “Class II or class II-like aldolase” proteins refers to proteins identified based on their similarity to bacterial 2,4-dihydroxy-hept-2-ene-1,7-dioic acid class II-like aldolase.

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 "DAG AT" or "DGAT" refers to a diacylglycerol acyltransferase (also known as an acyl-CoA-diacylglycerol acyltransferase or a diacylglycerol O-acyltransferase) (EC 2.3.1.20). This enzyme is responsible for the conversion of acyl-CoA and 1,2-diacylglycerol to TAG and CoA (thereby involved in the terminal step of TAG biosynthesis). Two families of DAG AT enzymes exist: DGAT1 and DGAT2. The former family shares homology with the acyl-CoA:cholesterol acyltransferase (ACAT) gene family, while the latter family is unrelated (Lardizabal et al., *J. Biol. Chem.* 276(42):38862-28869 (2001)).

The term "modulation" or "alteration" in the context of the present invention refers to increases or decreases of plastidic HpaIL aldolase 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.

The term “conserved domain” or “motif” means a set of amino acids conserved at specific positions along an aligned sequence of evolutionarily related proteins. While amino acids at other positions can vary between homologous proteins, amino acids that are highly conserved at specific positions indicate amino acids that are essential in the structure, the stability, or the activity of a protein. Because they are identified by their high degree of conservation in aligned sequences of a family of protein homologues, they can be used as identifiers, or “signatures”, to determine if a protein with a newly determined sequence belongs to a previously identified protein family.

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.

"Suppression DNA construct" is a recombinant DNA construct which when transformed or stably integrated into the genome of the plant, results in "silencing" of a target gene in the plant. The target gene may be endogenous or transgenic to the plant. "Silencing," as used herein with respect to the target gene, refers generally to the suppression of levels of mRNA or protein/enzyme expressed by the target gene, and/or the level of the enzyme activity or protein functionality. The terms "suppression", "suppressing" and "silencing", used interchangeably herein, include lowering, reducing, declining, decreasing, inhibiting, eliminating or preventing. "Silencing" or "gene silencing" does not specify mechanism and is inclusive, and not limited to, anti-sense, cosuppression, viral-suppression, hairpin suppression, stem-loop suppression, RNAi-based approaches, and small RNA-based approaches.

A suppression DNA construct may comprise a region derived from a target gene of interest and may comprise all or part of the nucleic acid sequence of the sense strand (or antisense strand) of the target gene of interest. Depending upon the approach to be utilized, the region may be 100% identical or less than 100% identical (e.g., at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 56%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 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%, or 99% identical) to all or part of the sense strand (or antisense strand) of the gene of interest.

Suppression DNA constructs are well-known in the art, are readily constructed once the target gene of interest is selected, and include, without limitation, cosuppression constructs, antisense constructs, viral-suppression constructs, hairpin suppression constructs, stem-loop suppression constructs, double-stranded RNA-producing constructs, and more generally, RNAi (RNA interference) constructs and small RNA constructs such as siRNA (short interfering RNA) constructs and miRNA (microRNA) constructs.

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

"Cosuppression" refers to the production of sense RNA transcripts capable of suppressing the expression of the target gene or gene product. "Sense" RNA refers to RNA transcript that includes the mRNA and can be translated into protein within a cell or *in vitro*. Cosuppression constructs in plants have been previously designed by focusing on overexpression of a nucleic acid sequence having homology to a native mRNA, in the sense orientation, which results in the reduction of all RNA having homology to the overexpressed sequence (see Vaucheret et al., *Plant J.* 16:651-659 (1998); and Gura, *Nature* 404:804-808 (2000)).

Another variation describes the use of plant viral sequences to direct the suppression of proximal mRNA encoding sequences (PCT Publication No. WO 98/36083 published on August 20, 1998).

Previously described is 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 No. WO 99/53050 published on October 21, 1999). In this case the stem is formed by polynucleotides corresponding to the gene of interest inserted in either sense or anti-sense orientation with respect to the promoter and the loop is formed by some polynucleotides of the gene of interest, which do not have a complement in the construct. This increases the frequency of cosuppression or silencing in the recovered transgenic plants. For review of hairpin suppression see Wesley, S.V. et al. (2003) Methods in Molecular Biology, Plant Functional Genomics: Methods and Protocols 236:273-286.

A construct where the stem is formed by at least 30 nucleotides from a gene to be suppressed and the loop is formed by a random nucleotide sequence has also effectively been used for suppression (PCT Publication No. WO 99/61632 published on December 2, 1999).

The use of poly-T and poly-A sequences to generate the stem in the stem-loop structure has also been described (PCT Publication No. WO 02/00894 published January 3, 2002).

Yet another variation includes using synthetic repeats to promote formation of a stem in the stem-loop structure. Transgenic organisms prepared with such recombinant DNA fragments have been shown to have reduced levels of the protein encoded by the nucleotide fragment forming the loop as described in PCT Publication No. WO 02/00904, published January 3, 2002.

RNA interference refers to the process of sequence-specific post-transcriptional gene silencing in animals mediated by short interfering RNAs (siRNAs) (Fire et al., *Nature* 391:806 (1998)). The corresponding process in plants is commonly referred to as post-transcriptional gene silencing (PTGS) or RNA silencing and is also referred to as quelling in fungi. The process of post-transcriptional gene silencing is thought to be an evolutionarily-conserved cellular defense mechanism used to prevent the expression of foreign genes and is commonly shared by diverse flora and phyla (Fire et al., *Trends Genet.* 15:358 (1999)). Such protection from foreign gene expression may have evolved in response to the production of double-stranded RNAs (dsRNAs) derived from viral infection or from the random integration of transposon elements into a host genome via a cellular response that specifically destroys homologous single-stranded RNA of viral genomic RNA. The presence of dsRNA in cells triggers the RNAi response through a mechanism that has yet to be fully characterized.

The presence of long dsRNAs in cells stimulates the activity of a ribonuclease III enzyme referred to as dicer. Dicer is involved in the processing of the dsRNA into short pieces of dsRNA known as short interfering RNAs (siRNAs) (Berstein et al., *Nature* 409:363 (2001)). Short interfering RNAs derived from dicer activity are typically about 21 to about 23 nucleotides in length and comprise about 19 base pair duplexes (Elbashir et al., *Genes Dev.* 15:188 (2001)). Dicer has also been implicated in the excision of 21- and 22-nucleotide small temporal RNAs (stRNAs) from precursor RNA of conserved structure that are implicated in translational control (Hutvagner et al., *Science* 293:834 (2001)). The RNAi response also features an endonuclease complex, commonly referred to as an RNA-induced silencing complex (RISC), which mediates cleavage of single-stranded RNA having sequence complementarity to the antisense strand of the siRNA duplex. Cleavage of the target RNA takes place in the middle of the region complementary to the antisense strand of the siRNA duplex. In addition, RNA

interference can also involve small RNA (e.g., miRNA) mediated gene silencing, presumably through cellular mechanisms that regulate chromatin structure and thereby prevent transcription of target gene sequences (see, e.g., Allshire, *Science* 297:1818-1819 (2002); Volpe et al., *Science* 297:1833-1837 (2002); Jenuwein, *Science* 297:2215-2218 (2002); and Hall et al., *Science* 297:2232-2237 (2002)). As such, miRNA molecules of the invention can be used to mediate gene silencing via interaction with RNA transcripts or alternately by interaction with particular gene sequences, wherein such interaction results in gene silencing either at the transcriptional or post-transcriptional level.

RNAi has been studied in a variety of systems. Fire et al. (*Nature* 391:806 (1998)) were the first to observe RNAi in *Caenorhabditis elegans*. Wianny and Goetz (*Nature Cell Biol.* 2:70 (1999)) describe RNAi mediated by dsRNA in mouse embryos. Hammond et al. (*Nature* 404:293 (2000)) describe RNAi in Drosophila cells transfected with dsRNA. Elbashir et al., (*Nature* 411:494 (2001)) describe RNAi induced by introduction of duplexes of synthetic 21-nucleotide RNAs in cultured mammalian cells including human embryonic kidney and HeLa cells.

Small RNAs play an important role in controlling gene expression. Regulation of many developmental processes, including flowering, is controlled by small RNAs. It is now possible to engineer changes in gene expression of plant genes by using transgenic constructs which produce small RNAs in the plant.

Small RNAs appear to function by base-pairing to complementary RNA or DNA target sequences. When bound to RNA, small RNAs trigger either RNA cleavage or translational inhibition of the target sequence. When bound to DNA target sequences, it is thought that small RNAs can mediate DNA methylation of the target sequence. The consequence of these events, regardless of the specific mechanism, is that gene expression is inhibited.

It is thought that sequence complementarity between small RNAs and their RNA targets helps to determine which mechanism, RNA cleavage or translational inhibition, is employed. It is believed that siRNAs, which are perfectly complementary with their targets, work by RNA cleavage. Some miRNAs have perfect or near-perfect complementarity with their targets, and RNA cleavage has been demonstrated for at least a few of these miRNAs. Other miRNAs have several mismatches with their targets, and apparently inhibit their targets at the translational

level. Again, without being held to a particular theory on the mechanism of action, a general rule is emerging that perfect or near-perfect complementarity causes RNA cleavage, whereas translational inhibition is favored when the miRNA/target duplex contains many mismatches. The apparent exception to this is microRNA 172 (miR172) in plants. One of the targets of miR172 is APETALA2 (AP2), and although miR172 shares near-perfect complementarity with AP2 it appears to cause translational inhibition of AP2 rather than RNA cleavage.

MicroRNAs (miRNAs) are noncoding RNAs of about 19 to about 24 nucleotides (nt) in length that have been identified in both animals and plants (Lagos-Quintana et al., *Science* 294:853-858 (2001), Lagos-Quintana et al., *Curr. Biol*. 12:735-739 (2002); Lau et al., *Science* 294:858-862 (2001); Lee and Ambros, *Science* 294:862-864 (2001); Llave et al., *Plant Cell* 14:1605-1619 (2002); Mourelatos et al., *Genes. Dev.* 16:720-728 (2002); Park et al., *Curr. Biol.* 12:1484-1495 (2002); Reinhart et al., *Genes. Dev.* 16:1616-1626 (2002)). They are processed from longer precursor transcripts that range in size from approximately 70 to 200 nt, and these precursor transcripts have the ability to form stable hairpin structures. In animals, the enzyme involved in processing miRNA precursors is called dicer, an RNAse III-like protein (Grishok et al., *Cell* 106:23-34 (2001); Hutvagner et al., *Science* 293:834-838 (2001); Ketting et al., *Genes. Dev.* 15:2654-2659 (2001)). Plants also have a dicer-like enzyme, DCL1 (previously named CARPEL FACTORY/SHORT INTEGUMENTS1/ SUSPENSOR1), and recent evidence indicates that it, like dicer, is involved in processing the hairpin precursors to generate mature miRNAs (Park et al., *Curr. Biol.* 12:1484-1495 (2002); Reinhart et al., *Genes Dev.* 16:1616-1626 (2002)). Furthermore, it is becoming clear from recent work that at least some miRNA hairpin precursors originate as longer polyadenylated transcripts, and several different miRNAs and associated hairpins can be present in a single transcript (Lagos-Quintana et al., *Science* 294:853-858 (2001); Lee et al., *EMBO J.* 21:4663-4670 (2002)). Recent work has also examined the selection of the miRNA strand from the dsRNA product arising from processing of the hairpin by DICER (Schwartz et al., *Cell* 115:199-208 (2003)). It appears that the stability (i.e. G:C versus A:U content, and/or mismatches) of the two ends of the processed dsRNA affects the strand selection, with the low stability end being easier

to unwind by a helicase activity. The 5' end strand at the low stability end is incorporated into the RISC complex, while the other strand is degraded.

MicroRNAs (miRNAs) appear to regulate target genes by binding to complementary sequences located in the transcripts produced by these genes. In the case of lin-4 and let-7, the target sites are located in the 3' UTRs of the target mRNAs (Lee et al., *Cell* 75:843-854 (1993); Wightman et al., *Cell* 75:855-862 (1993); Reinhart et al., *Nature* 403:901-906 (2000); Slack et al., *Mol. Cell* 5:659-669 (2000)), and there are several mismatches between the lin-4 and let-7 miRNAs and their target sites. Binding of the lin-4 or let-7 miRNA appears to cause downregulation of steady-state levels of the protein encoded by the target mRNA without affecting the transcript itself (Olsen and Ambros, *Dev. Biol*. 216:671-680 (1999)). On the other hand, recent evidence suggests that miRNAs can in some cases cause specific RNA cleavage of the target transcript within the target site, and this cleavage step appears to require 100% complementarity between the miRNA and the target transcript (Hutvagner and Zamore, *Science* 297:2056-2060 (2002); Llave et al., *Plant Cell* 14:1605-1619 (2002)). It seems likely that miRNAs can enter at least two pathways of target gene regulation: (1) protein downregulation when target complementarity is <100%; and (2) RNA cleavage when target complementarity is 100%. MicroRNAs entering the RNA cleavage pathway are analogous to the 21-25 nt short interfering RNAs (siRNAs) generated during RNA interference (RNAi) in animals and posttranscriptional gene silencing (PTGS) in plants, and likely are incorporated into an RNA-induced silencing complex (RISC) that is similar or identical to that seen for RNAi.

Identifying the targets of miRNAs with bioinformatics has not been successful in animals, and this is probably due to the fact that animal miRNAs have a low degree of complementarity with their targets. On the other hand, bioinformatic approaches have been successfully used to predict targets for plant miRNAs (Llave et al., *Plant Cell* 14:1605-1619 (2002); Park et al., *Curr. Biol.* 12:1484-1495 (2002); Rhoades et al., *Cell* 110:513-520 (2002)), and thus it appears that plant miRNAs have higher overall complementarity with their putative targets than do animal miRNAs. Most of these predicted target transcripts of plant miRNAs encode members of transcription factor families implicated in plant developmental patterning or cell differentiation.

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 do 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 effect 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 85% 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 thatcodes 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).

"Mature" protein refers to a post-translationally processed polypeptide (i.e., one from which any pre- or propeptides 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 propeptides still present). Pre- and propeptides may be but are not limited to intracellular localization signals.

A "signal peptide" is an amino acid sequence that is translated in conjunction with a protein and directs the protein to the secretory system (Chrispeels, M. (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol. 42*:21-53). If the protein is to be directed to a vacuole, a vacuolar targeting signal (supra) can further be added, or if to the endoplasmic reticulum, an endoplasmic reticulum retention signal (supra) may be added. If the protein is to be directed to the nucleus, any signal peptide present should be removed and instead a nuclear localization signal included (Raikhel, N. (1992) *Plant Phys. 100*:1627-1632). A "chloroplast transit peptide" is an amino acid sequence that is translated in conjunction with a protein and directs the protein to the chloroplast or other plastid types present in the cell in which the protein is made. "Chloroplast transit sequence" refers to a nucleotide sequence that encodes a chloroplast transit peptide.

Chloroplast transit sequences can be predicted for example by using ChloroP at the online ChloroP 1.1. Server, which predicts the presence of chloroplast transit peptides (cTP) in protein sequences and the location of potential cTP cleavage sites.

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 plastdic HpaIL aldolase polypeptides described herein in plants. The size of the oil, protein, starch and soluble carbohydrate pools in soybean seeds as well as the seed weight can be modulated or altered (i.e. increased or decreased) by altering the expression of a specific gene, plastidic Hpal aldolase polypeptides.

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: 29, 31, 33, 35, 49, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, and 147and wherein seed obtained from said transgenic plant has an altered oil, protein, starch and/or soluble carbohydrate content and/or altered seed weight 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: 29, 31, 33, 35, 49, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, and 147and wherein said transgenic seed has an altered oil, protein, starch and/or soluble carbohydrate content and/or altered seed weight 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: 29, 31, 33, 35, 49, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, and 147 and wherein said transgenic seed has an increased protein 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 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: 29, 31, 33, 35, 49, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, and 147and 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: 29, 31, 33, 35, 49, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, and 123; 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 HpaIL aldolase 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: 28, 30, 32, 34, 48, 124, 125, 126,127, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, or 146; 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 HpaIL aldolase.

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 HpaIL aldolase 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 HpaIL aldolase RNA, protein or enzyme activity in the host cell containing the isolated polynucloetide or recombinant DNA construct; (d) comparing the level of the HpaIL aldolase RNA, protein or enzyme activity in the host cell containing the isolated poynucleotide or recombinant DNA construct with the level of the HpaIL aldolase 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 HpaIL aldolase 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 plastidic HpaIL aldolase 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 plastidic polypeptide having HpaIL aldolase activity has been suppressed.

Preferably, the gene encodes a plastidic polypeptide having Hpal aldolase 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: 28, 30, 32, 34, 48, 124, 125, 126,127, 133, 134, 135, 136, 137, 138,

139, 140, 141, 142, 143, 144, 145, or 146, or (ii) the complement of (i); wherein (i) or (ii) is useful in co-suppression or antisense suppression of endogenous HpaIL aldolase 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: 29, 31, 33, 35, 49, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, and 147or (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: 29, 31, 33, 35, 49, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, and 123, 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 plastidic HpaIL aldolase, and wherein said plant has an altered oil, protein, starch and/or soluble carbohydrate content and/or altered seed weight, 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: 28, 30, 32, 34, 48, 124, 125, 126,127, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, or 146; or (b) the full-length complement of (a):wherein (a) or (b) is of sufficient length to inhibit expression of endogenous plastidic HpaIL aldolase 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: 28, 30, 32, 34, 48, 124, 125, 126,127, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, or 146; or (b) the full-length complement of (a):
wherein (a) or (b) is of sufficient length to inhibit expression of plastidic HpaIL aldolase 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: 29, 31, 33, 35, 49, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, and 123; and (b) regenerating a plant from the transformed plant cell.

The method may further comprise (c) obtaining a progeny plant derived from the transgenic plant, wherein said progeny plant comprises in its genome the recombinant DNA construct and exhibits an altered oil, protein, starch and/or soluble carbohydrate content and/or altered seed weight, when compared to a control plant not comprising the recombinant DNA construct.

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: 29, 31, 33, 35, 49, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, and 123; 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 and/or altered seed weight, 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: 29, 31, 33, 35, 49, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, and 123; 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 protein 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: 28, 30, 32, 34, 48, 124, 125, 126,127, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, or 146; or (ii) the full-length complement of (i); wherein (i) or (ii) is of sufficient length to inhibit expression of plastidic HpaIL aldolase 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 and/or altered seed weight, 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: 28, 30, 32, 34, 48, 124, 125, 126,127, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, or 146; or (ii) the full-length complement of (i);
wherein (i) or (ii) is of sufficient length to inhibit expression of plastidic HpaIL aldolase 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.

A transgenic oilseed of the invention can comprise at least one DGAT sequence and a construct downregulating plastidic Hpal or Hpal-like activity, wherein the DGAT sequence and the plastidic Hpal or Hpal-like construct can be in the same recombinant construct or in separate recombinant constructs, and wherein seed obtained from said transgenic plant has an increased oil content when compared to the oil content of seed obtained from a control plant not comprising said construct or when compared to transgenic seed obtained from a transgenic plant comprising either said DGAT sequences alone or said construct

downregulating HpAI or HpaI-like activity alone.

Those skilled in the art will appreciate that the instant invention includes, but is not limited to, the DGAT and plastidic HpaI and HpaI-like sequences disclosed herein. For example, the DGAT sequence can be selected from the group consisting of DGAT1, DGAT2 and DGAT1 in combination with DGAT2.

In one embodiment the present invention concerns a transgenic plant comprising at least one DGAT sequence and a construct downregulating plastidic HpaI or HpaI-like activity, wherein the DGAT sequence and the plastidic HpaI or HpaI-like construct can be in the same recombinant construct or in separate recombinant constructs, and wherein seed obtained from said transgenic plant has an increased oil content when compared to the oil content of seed obtained from a control plant not comprising said construct or when compared to transgenic seed obtained from a transgenic plant comprising either said DGAT sequences alone or said construct downregulating HpaI-like activity alone. Such increases in the oil content would include, but are not limited to, at least 0,5%, 0.6%, 0.7%, 0.8%, 0.9%, 1%, 1.1.%, 1.2%, 1.3%, 1.4%, 1.5%, 1.6%, 1.7%, 1.8%, 1.9%, 2%, 2.1.%, 2.2%, 2.3%, 2.4%, 2.5%, 2.6%, 2.7%, 2.8%, 2.9%, 3%, 3.1.%, 3.2%, 3.3%, 3.4%, 3.5%, 3.6%, 3.7%, 3.8%, 3.9%, 4%, 4.1.%, 4.2%, 4.3%, 4.4%, 4.5%, 4.6%, 4.7%, 4.8%, 4.9%, 5%, 5.1.%, 5.2%, 5.3%, 5.4%, 5.5%, 5.6%, 5.7%, 5.8%, 5.9%, 6%, 6.1%, 6.2%, 6.3%, 6.4%, 6.5%, 6.6%, 6.7%, 6.8%, 6.9%, 7%, 7.1%, 7.2%, 7.3%, 7.4%, 7.5%, 7.6%, 7.7%, 7.8%, 7.9%,8%, 8.1.%, 8.2%, 8.3%, 8.4%, 8.5%, 8.6%, 8.7%, 8.8%, 8.9%, 9%, 9.1%, 9.2%, 9.3%, 9.4%, 9.5%, 9.6%, 9.7%, 9.8%, 9.9%, 10%, 10.1%, 10.2%, 10.3%, 10.4%, 10.5%, 10.6%, 10.7%, 10.8%, 10.9%,11%, 11.1%, 11.2%, 11.3%, 11.4%, 11.5%, 11.6%, 11.7%, 11.8%, 11.9%,12%, 12.1%, 12.2%, 12.3%, 12.4%, 12.5%, 12.6%, 12.7%, 12.8%, 12.9%,13%, 13.1%, 13.2%, 13.3%, 13.4%, 13.5%, 13.6%, 13.7%, 13.8%, 13.9%,14%, 14.1%, 14.2%, 14.3%, 14.4%, 14.5%, 14.6%, 14.7%, 14.8%, 14.9%, 15%, 15.1%, 15.2%, 15.3%, 15.4%, 15.5%, 15.6%, 15.7%, 15.8%, 15.9%, 16%, 16.1%, 16.2%, 16.3%, 16.4%, 16.5%, 16.6%, 16.7%, 16.8%, 16.9%, 17%, 17.1%, 17.2%, 17.3%, 17.4%, 17.5%, 17.6%, 17.7%, 17.8%, 17.9%, 18%, 18.1%, 18.2%, 18.3%, 18.4%, 18.5%, 18.6%, 18.7%, 18.8%, 18.9%, 19%, 19.1%, 19.2%, 19.3%, 19.4%, 19.5%, 19.6%, 19.7%, 19.8%, 19.9%, 20%, 20.1%, 20.2%, 20.3%, 20.4%, 20.5%, 20.6%, 20.7%, 20.8%, 20.9%, 21%, 21.1%, 21.2%, 21.3%, 21.4%, 21.5%, 21.6%, 21.7%, 21.8%, 21.9%, 22%, 22.1%,

22.2%, 22.3%, 22.4%, 22.5%, 22.6%, 22.7%, 22.8%, 22.9%, 23%, 23.1%, 23.2%, 23.3%, 23.4%, 23.5%, 23.6%, 23.7%, 23.8%, 23.9%, 24%, 24.1%, 24.2%, 24.3%, 24.4%, 24.5%, 24.6%, 24.7%, 24.8%, 24.9%, 25%, 25.1%, 25.2%, 25.3%, 25.4%, 25.5%, 25.6%, 25.7%, 25.8%, 25.9%, 26%, 26.1%, 26.2%, 26.3%, 26.4%, 26.5%, 26.6%, 26.7%, 26.8%, 26.9%, 27%, 27.1%, 27.2%, 27.3%, 27.4%, 27.5%, 27.6%, 27.7%, 27.8%, 27.9%, 28%, 28.1%, 28.2%, 28.3%, 28.4%, 28.5%, 28.6%, 28.7%, 28.8%, 28.9%, 29%, 29.1%, 29.2%, 29.3%, 29.4%, 29.5%, 29.6%, 29.7%, 29.8%, 29.9%, or 30 %, on a dry-weight basis.

Further embodiments include transgenic seed obtained from the transgenic plant of claim1 comprising at least one DGAT sequence and a construct downregulating HpaI or HpaI-like activity, wherein the DGAT sequence and the plastidic HpaI-like construct can be in the same recombinant construct or in separate recombinant constructs and wherein the oil content of said transgenic seed is increased when compared to the oil content of control seed not comprising said construct or transgenic seed comprising either said DGAT sequence alone or said construct downregulating HpAI or HpaI-like activity alone.

Transgenic seed obtained from a monocot or dicot plant are included in the invention, e.g. maize or soybean.

Another embodiment of the present invention comprises a seed-specific or seed-preferred promoter as the at least one regulatory element linked to the nucleic acid sequences of the present invention. Also, endosperm or embryo-specific promoter are included.

Another embodiment of the present invention comprises q method for increasing the oil content of a seed comprising: a) transforming at least one cell with at least one recombinant construct having at least one DGAT sequence and a construct downregulating plastidic HpaI or HpaI-like activity wherein the DGAT sequence and the HpaI or HpaI-like construct can be in the same recombinant construct or in separate recombinant constructs; (b) selecting the transformed soybean cell(s) of step (a) having an increased oil content when compared to the oil content of a control cell not comprising said construct or when compared to to transgenic seed obtained from a transgenic plant comprising either said DGAT sequences alone or said construct downregulating HpAI or HpaI-like activity alone.

Product and/or by-product obtained from the transgenic seed of transformed with any of the recombinant construct of the present invention are also included.

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, or a control not comprising the recombinant construct 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) *BioITechnology* 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.

Typically, when a transgenic plant comprising a recombinant DNA construct or suppression DNA construct in its genome exhibits an altered , e.g. increased/ or decreased oil, protein, soluble carbohydrate or starch content relative to a reference or control plant, the reference or control plant does not comprise in its genome the recombinant DNA construct or suppression DNA construct.

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 plastidic HpaIL aldolase, most preferably in antisense inhibition or cosuppression of an plastidic HpaIL aldolase 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.

One embodiment of the invention comprises an isolated polynucleotide comprising:(a) a nucleotide sequence encoding a polypeptide with HpAIL aldolase activity, wherein, based on the Clustal V method of alignment with pairwise alignment default parameters of KTUPLE=1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5, the polypeptide has an amino acid sequence of at least 75% sequence identity when compared to SEQ ID NO:120, 121, 122 or 123, or (b) the full complement of the nucleotide sequence of (a).

Furthermore, the amino acid sequence of the polypeptide can comprise SEQ ID NO: 120, 121, 122 or 123.

The present invention further comprises an isolated polynucleotide encoding a polypeptide, wherein said polynucleotide is capable of altering the endogenous expression of plastidic HpaI-like activity and wherein said polypeptide comprises a chloroplast transit peptide and at least one motif selected from the group consisting of: SEQ ID NO: 128, 129, 130, 131, or 132.

Another embodiment includes an isolated polynucleotide encoding a plant HpaI-like polypeptide, wherein said polynucleotide is capable of altering the endogenous expression of plastidial HpaI-like activity and wherein said polypeptide has a Km (acetaldehyde) at least 1.7 fold lower than than the Km (acetaldehyde) of bacterial HpaIL aldolase activity and a Vmax of at least 15 fold lower than the bacterial HpaIL aldolase activity. Useful Km values of plastidial HpaI-like activity are at least 1.7, 1.75, 1.8, 1.85, 1.9, 1.95, 2.0, or 2.5 fold lower compared to the

bacterial HpaII aldolase activity, inparticular the aldolase of *P. putida.* Useful Vmax values for plastidial Hpal-like activity are at least 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59,or 60 fold lower compared to the bacterial HpaII aldolase activity, inparticular the aldolase of *P. putida.*

In another embodiment, the present invention includes a vector comprising any of the isolated polynucleotides of the present invention.

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

In another embodiment, the present invention includes a method for producing a transgenic plant comprising transforming a plant cell with any of the isolated polynucleotides or recombinant DNA constructs of the present invention and regenerating a transgenic plant from the transformed plant cell. The invention is also directed to the transgenic plant produced by this method, and transgenic seed obtained from this transgenic plant.

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. 1), 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 lo22048

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 500 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 or soybean somatic embryos was calculated as follows:

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

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 K22048 is only 90 % of that of WT control plants (pooled seed of six WT plants) grown in the same flat.

TABLE 4

Oil Content of T4 activation-tagged lines derived from T3 pool 500

BARCODE	% Oil	T3 pool ID #	oil content % of WT
K22048	**33.6**	**500**	**90**
K22049	41.6	500	111.3
K22050	38.7	500	103.5
K22051	41	500	109.8
K22052	38.7	500	103.5
K22053	41	500	109.6
K22054	38.8	500	103.8
K22055	41.7	500	111.5
K22056	40	500	107
K22057	39.8	500	106.4
K22058	39.4	500	105.4
K22059	34.4	500	92.1
K22060	39.8	500	106.4
K22061	37.6	500	100.6
K22062	40.4	500	108.1
K22063	37.9	500	101.3
K22064	39.8	500	106.4
K22065	41	500	109.7
K22066	41.2	500	110.2
K22067	39.7	500	106.3

K22068	37.7	500	100.8
K22069	36.4	500	97.4
K22070	38.1	500	102
K22071	40.9	500	109.3
K22072	41.3	500	110.4
K22073	40.1	500	107.4
K22074	35.7	500	95.6
K22075	39.3	500	105.2
K22076	38.4	500	102.8
wt	37.4		

K22048 was renamed lo22048. 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 lo22048

BARCODE	% Oil	Average % oil	T5 activation-tagged line ID	oil content % of WT	Average oil content % of WT
	37.3		lo22048	103.5	
	35.6		lo22048	98.9	
	34.5		lo22048	96.0	
	34.5		lo22048	95.8	
	34.4		lo22048	95.6	
	34.4		lo22048	95.5	
	33.8		lo22048	93.8	
	33.5		lo22048	93.1	
	33.4		lo22048	92.7	
	32.8	34.4	lo22048	91.1	95.6
	37.1		WT		
	36.7		WT		
	35.9		WT		
	35.8		WT		
	34.6	36	WT		

Table 5 shows that the seed oil content of T5 activation-tagged line lo22048 is between 91.1 and 103.5 % of that of WT control plants grown in the same flat. The average oil content of all T5 lines of lo22048 was 95.6 % of the WT control plants. 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 lo22048-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 lo22048. Twenty-four Basta-resistant T5 seedling of lo22048 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 lo22048 and 12 WT plants. Oil content of lo22048 and WT seed was measured by NMR (Table 6).

TABLE 6

Oil Content of T6 activation-tagged line lo22048

Barcode	% Oil	Seed ID	oil content % of WT
K35910	40.1	lo22048	90.7
K35911	44.2	WT	

T6 seed of lo22048 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 ºC) 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 ºC. After injection (2 µl, 20:1 split) the initial column temperature (150 ºC) was increased to 180 ºC at a rate of 3 ºC/min and then at 25 ºC/min to a final temperature of 320 ºC. 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 lo22048 and WT Control Seed

Genotype	Bar code ID	Oil (%, NMR)	Protein %	Seed Weight (µg)	fructose (µg mg^{-1} seed)
lo22048	K35910	40.1	16.3	26.0	0.65
WT	K35911	44.2	15.22	23.7	0.59
	Δ TG/WT %	-9.2	+7.1	+10	+10.1
Genotype	Bar code ID	glucose (µg mg^{-1} seed)	sucrose (µg mg^{-1} seed)	raffinose (µg mg^{-1} seed)	stachyose (µg mg^{-1} seed)

lo22048	K35910	9.17	28.13	0.61	3.3
WT	K35911	7.45	26.25	0.52	2.9
	ΔTG/WT %	+23.1	+7.1	+17.3	+13.8

The oil decrease in seed oil content of lo22048 is associated with an increase in seed weight and protein. The soluble carbohydrate profile of lo22048 differs from that of WT seed. The former shows increase in soluble carbohydrates including fructose, glucose, raffinose and stachyose. Herbicide-resistant seedlings were grown in soil. Pollen of lo22048 plant was used to fertilize emasculated immature flowers of WT plants. F1 seed were germinated on selective media, transferred to soil and 10 herbicide-resistant F1 plants were grown alongside four WT plants and four lo22048 parent plants in the same flat. Parent seed were bulk harvested. F2 seed of lo22048 were harvested from individual plants. Table 8 shows that 8 out of 10 F_1 plants produced seed with an oil content that was lower than that of WT seed grown in the same flat. The average decrease in seed oil content (compared to WT) of all F1 plants was 92.4 % which is very close to 94.2 % which was observed for the lo22048 parent.

TABLE 8

Seed oil content of F1 plants derived from a cross of lo22048 to WT plants of ecotype Columbia

genotype	BARCODE	% oil	oil content % of wt	avg. oil content % of WT
lo22048xCOL F_1	K41190	41.9	105.3	
	K41188	39.8	100.0	
	K41187	37.8	95.1	
	K41195	37.7	94.8	
	K41186	37.1	93.2	
	K41189	36.8	92.6	
	K41191	35.3	88.8	
	K41192	35.3	88.7	
	K41194	33.1	83.3	
	K41193	32.9	82.7	92.4
lo22048	K41196	37.5	94.3	
wt	K41197	39.8		

In summary the lo22048 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 5-10 % compared to WT) that is accompanied by increased seed size, protein content and increased levels of soluble carbohydrate in mature dry seed.

EXAMPLE 3

Identification of Activation-Tagged Genes

Genes flanking the T-DNA insert in the lo22048 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 lo22048

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 (lo22048) 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 1347–1543 of the Arabidopsis gene At4g10760. lo22048 carries a T-DNA insertion in the first intron of At4g10760 which very likely disrupts the function of this gene. Disruption of this gene is known to result in an embryo defective phenotype characterized by developmental arrest at the globular stage. (Zhong S. et al Plant Cell (2008), 20, 1278-1288). Because of the location of the T-DNA in lo22048 we conclude that like the emb1706 alleles of At4g10760 the lo22048-T DNA insertion allele of At4g10760 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 lo22048 transgene suggests that the disruption of At4g10760 is not related to the low seed oil phenotype of lo22048.

Validation of Candidate *Arabidopsis* Gene (At4g10750) via Transformation into *Arabidopsis*

The gene At4g10750, specifically its inferred start codon is 3.25 kb upstream of the SAIFF sequence corresponding to sequence adjacent to the left T-DNA border in lo22048. This gene is annotated as encoding a possibly plastidic, soluble protein with similarity to bacterial 2,4-dihydroxy-hept-2-ene-1,7-dioic acid and is subsequently called HpaI-like (HpaIL).

Primers HpaILORF FWD (SEQ ID NO:11) and HpaIL ORF REV (SEQ ID NO:12) were used to amplify the At4g10750 ORF from genomic DNA of Arabidopsis plants of the Columbia ecotype. The PCR product was cloned into pENTR (Invitrogen, USA) to give pENTR-HpaIL (SEQ ID NO:13). The HpaIL 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-HpaIL is set forth as SEQ ID NO:14.

pKR1478-HpaIL (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- HpaIL 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. T1 Plants are grown to maturity alongside wt contral plants and T2 seeds are harvested.

EXAMPLE 5

Seed-specific RNAi of At4g10750. 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 HpaIL UTR FWD (SEQ ID NO:24) and HpaIL UTR REV (SEQ ID NO:25) were used to amplify the At4g10750 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-HpaIL 3'UTR (SEQ ID NO:26).

5 μg of plasmid DNA of pENTR-HpaIL1 3'UTR (SEQ ID NO:26) was digested with EcoRV/HpaI. A restriction fragment of 528 bp (derived from pENTR-HpaIL1 3'UTR) was excised from an agarose gel. Purified gene fragments of the 3'UTR sequence were inserted into vector pKR1482 using LR clonase (Invitrogen) according to the manufacturers instructions, to give pKR1482 HpaIL 3'UTR (SEQ ID NO:27)

pKR1482 HpaIL 3'UTR (SEQ ID NO:27) 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 pKR1482 HpaIL 3'UTR (SEQ ID NO:27) 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 16 events were generated with pKR1482 HpaIL. Four wild-type (WT) control plants were grown in the same flat. WT seeds were bulk harvested and T2 seeds of individual 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 vector pKR1482-HpaIL 3'UTR for seed specific gene suppression of At4g10750

Construct	BARCODE	% oil	oil content % of WT	avg. oil content % of WT
pKR1482 HpaIL 3'UTR	K14724	42.2	107.4	
	K14729	41.6	106.0	
	K14734	**41.6**	**105.9**	
	K14733	**41.6**	**105.9**	

	K14719	41.4	105.5	
	K14732	41.4	105.4	
	K14727	41.4	105.4	
	K14721	41.0	104.4	
	K14730	40.5	103.2	
	K14728	40.4	102.9	
	K14725	40.4	102.7	
	K14723	38.1	97.0	
	K14731	38.1	97.0	
	K14720	38.0	96.7	
	K14726	37.3	94.9	
	K14735	35.3	89.9	101.9
wt	K14736	39.6		
	K14737	39.0		

Table 10 shows that seed-specific down regulation of At4g10750 leads to increased oil content in Arabidopsis seed.

T2 seed of events K14733 and K14734 that both carry transgenes pKR1482 HpaIL 3'UTR were plated on plant growth media containing kanamycin. For event K14733 and K14734 21 and 23 kanamycin-resistant T2 seedlings, respectively, were grown to maturity alongside WT plants of the Columbia ecotype grown in the same flats. Oil content of T3 seed is depicted in Table 11. Table 11 demonstrates that the oil increase associated with seed-specific down regulation of At4g10750 is heritable.

TABLE 11

Seed oil content of T2 plants generated with binary vectors pKR1482-HpaIL 3'UTR for seed specific gene suppression of At4g10750

Construct	**Event**	**T2 plant #**	**%oil**	**Oil content % of wt**	**Avg. oil content % of wt**
pKR1482 HpaIL 3'UTR	**K14733**	1	44.6	109.0	
		2	44.3	108.2	
		3	44.2	107.8	
		4	43.9	107.2	
		5	43.4	105.9	
		6	43.4	105.9	
		7	43.3	105.8	
		8	42.2	103.1	
		9	42.1	102.8	
		10	42.0	102.5	

		11	42.0	102.5	
		12	41.9	102.4	
		13	41.9	102.4	
		14	41.9	102.3	
		15	41.4	101.0	
		16	41.2	100.5	
		17	41.1	100.4	
		18	40.9	99.8	
		19	40.0	97.7	
		20	39.9	97.4	
		21	39.7	96.9	102.9
Wt		1	42.8		
		2	42.6		
		3	42.4		
		4	42.3		
		5	41.9		
		6	41.6		
		7	41.3		
		8	40.9		
		9	40.3		
		10	39.8		
		11	38.2		
		12	37.4		
pKR1482 HpaIL 3'UTR	**K14734**	1	43.5		
		2	43.3	113.2	
		3	43.2	112.7	
		4	43.0	112.5	
		5	42.9	111.9	
		6	42.8	111.5	
		7	42.8	111.4	
		8	42.7	111.3	
		9	42.0	111.0	
		10	41.7	109.2	
		11	41.4	108.6	
		12	41.2	107.6	
		13	41.2	107.3	
		14	40.7	107.2	
		15	40.7	105.8	
		16	40.7	105.8	
		17	40.2	105.8	
		18	39.9	104.7	
		19	39.8	103.9	
		20	38.9	103.4	
		21	38.6	101.3	
		22	37.4	100.5	
		23	36.3	97.3	106.9
Wt		1	39.9		
		2	39.8		

3	39.6
4	39.3
5	39.2
6	38.8
7	38.2
8	37.6
9	37.4
10	36.7
11	36.5

EXAMPLE 6

Identification of genes of *Brassica napus* closely-related to At4g10750

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 At4g10750 and tBLASTn. The assembly encompasses about 558465 public ESTs and has a total of 90310 sequences (47591 assemblies and 42719 singletons). There is one gene which shares 84.5 % amino acid sequence identity to At4g10750. This genes, its % identity to At4g10750 and SEQ ID NOs are listed in Table 12.

TABLE 12

Brassica napus gene closely related to At4g10750

Gene name	% AA sequence identity to At4g10750	SEQ ID NO: NT	SEQ ID NO: AA
TC 25873	**84.5**	**28**	**29**

EXAMPLE 7

Identification of genes of soybean (*Glycine max*) closely-related to At4g10750

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 At4g10750 and tBLASTn. There is one gene which shares 61.3 % amino acid sequence identity At4g10750. This gene, its properties and SEQ ID NO is listed in Table 13

TABLE 13

Soybean gene closely related to At4g10750

Gene name	% AA sequence identity to At4g10750	SEQ ID NO: NT	SEQ ID NO: AA
Glyma09g21760	61.3	30	31

EXAMPLE 8

Identification of genes of maize (*Zea mays*) closely-related to At4g10750

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 At4g10750 and tBLASTn. There is one gene which shares 56.3 % amino acid sequence identity to At4g10750, its properties and SEQ ID NOs are listed in Table 14

TABLE 14

Maize gene closely related to At4g10750

Gene name	% AA sequence identity to At4g10750	SEQ ID NO: NT	SEQ ID NO: AA
PCO651314	56.3	32	33

EXAMPLE 9

Identification of genes of rice (*Oryza sativa*) closely-related to At4g10750

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 At4g10750 and tBLASTn. There is gene which share at least 56.4% amino acid sequence identity to At4g10750. This gene, its properties and SEQ ID NOs is listed in Table 15.

TABLE 15

Rice genes closely related to At4g10750

Gene name	% AA sequence identity to At4g10750	SEQ ID NO: NT	SEQ ID NO: AA
Os09g36030	56.4	34	35

EXAMPLE 10

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 11

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 12

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 13

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 14

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:36

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 plastidic HpaIL aldolase gene expression, protein or enzyme activity in a plant cell.

EXAMPLE 15

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:37) 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:38

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:39

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:39) into the NotI site of pKS67. The ELVISLIVES fragment is made by PCR using two primers and no other DNA:

SEQ ID NO:40
5'-
GAATTCCGGCCGGAGCTGGTCATCTCGCTCATCGTCGAGTCGGCGGCCGCC
GACTCGACGATGAGCGAGATGACCAGCTCCGGCCGGAATTC-3'

SEQ ID NO:41
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 16

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 plastidic HpaIL aldolase 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 plastidic HpaIL aldolase 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 17

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 plastidic HpaIL aldolase expression is expected to lead to an increase in seed oil content. In contrast overexpression of HpaIL aldolase in the endosperm-specific is expected to lead to an increase in seed starch content.

EXAMPLE 18

Seed specifc RNAi of HpaIL in soybean

A plasmid vector (pKS423) for generation of transgenic soybean events that show seed specific down-regulation of the soy HpaIL (Glyma09g21760) gene was constructed.

Briefly plasmid DNA of applicants EST clone sfp1n.pk034.b9 corrresponding to Glyma09g21760 (SEQ ID NO:30) was used in two PCR reactions with either Primers SA64 (SEQ ID NO:42) and SA65 (SEQ ID NO:43) or SA66 (SEQ ID NO:44) and SA64 (Seq ID NO:42). 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 SA64 PCR primer. A PCR product of approximately 1 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 0.968 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 pKS423 (SEQ ID NO:45).

Plasmid DNA of pKS423 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 pKS423 determined as described in Example 17.

The plasmid vector pKS123 is described in PCT Application No. WO 02/08269. Plasmid pKS120 (SEQ ID NO: 46) is identical to pKS123 (*supra*) with the exception that the *Hind*III fragment containing Bcon/*Not*I/Phas3' cassette was removed.

Generation of transgenic somatic embryos:

Soybean somatic embryos soybean tissue was co-bombarded as described below with a plasmid DNA of pKS120 or pKS423.

Culture Conditions:

Soybean embryogenic suspension cultures (cv. Jack) were maintained in 35 mL liquid medium SB196 (*infra*) on a rotary shaker, 150 rpm, 26 °C with cool white fluorescent lights on 16:8 h day/night photoperiod at light intensity of 60-85 µE/m2/s. Cultures were subcultured every 7 days to two weeks by inoculating approximately

35 mg of tissue into 35 mL of fresh liquid SB196 (the preferred subculture interval is every 7 days).

Soybean embryogenic suspension cultures were transformed with the soybean expression plasmids by the method of particle gun bombardment (Klein et al., *Nature* 327:70 (1987)) using a DuPont Biolistic PDS1000/HE instrument (helium retrofit) for all transformations.

Soybean Embryogenic Suspension Culture Initiation:

Soybean cultures were initiated twice each month with 5-7 days between each initiation. Pods with immature seeds from available soybean plants 45-55 days after planting were picked, removed from their shells and placed into a sterilized magenta box. The soybean seeds were sterilized by shaking them for 15 min in a 5% Clorox solution with 1 drop of ivory soap (i.e., 95 mL of autoclaved distilled water plus 5 mL Clorox and 1 drop of soap, mixed well). Seeds were rinsed using 2 1-liter bottles of sterile distilled water and those less than 4 mm were placed on individual microscope slides. The small end of the seed was cut and the cotyledons pressed out of the seed coat. Cotyledons were transferred to plates containing SB199 medium (25-30 cotyledons per plate) for 2 weeks, then transferred to SB1 for 2-4 weeks. Plates were wrapped with fiber tape. After this time, secondary embryos were cut and placed into SB196 liquid media for 7 days.

Preparation of DNA for Bombardment:

Plasmid DNA of pKS120 or pKS423 were used for bombardment.

A 50 µL aliquot of sterile distilled water containing 1 mg of gold particles was added to 5 µL of a 1 µg/µL plasmid DNA solution 50 µL 2.5M $CaCl_2$ and 20 µL of 0.1 M spermidine. The mixture was pulsed 5 times on level 4 of a vortex shaker and spun for 5 sec in a bench microfuge. After a wash with 150 µL of 100% ethanol, the pellet was suspended by sonication in 85 µL of 100% ethanol. Five µL of DNA suspension was dispensed to each flying disk of the Biolistic PDS1000/HE instrument disk. Each 5 µL aliquot contained approximately 0.058 mg gold particles per bombardment (i.e., per disk).

Tissue Preparation and Bombardment with DNA:

Approximately 100-150 mg of 7 day old embryonic suspension cultures were placed in an empty, sterile 60 x 15 mm petri dish and the dish was placed inside of an empty 150 x 25 mm Petri dish. Tissue was bombarded 1 shot per plate with

membrane rupture pressure set at 650 PSI and the chamber was evacuated to a vacuum of 27-28 inches of mercury. Tissue was placed approximately 2.5 inches from the retaining /stopping screen.

Selection of Transformed Embryos:

Transformed embryos were selected using hygromycin as the selectable marker. Specifically, following bombardment, the tissue was placed into fresh SB196 media and cultured as described above. Six to eight days post-bombardment, the SB196 is exchanged with fresh SB196 containing 30 mg/L hygromycin. The selection media was refreshed weekly. Four to six weeks post-selection, green, transformed tissue was observed growing from untransformed, necrotic embryogenic clusters. Isolated, green tissue was removed and inoculated into multi-well plates to generate new, clonally propagated, transformed embryogenic suspension cultures.

Embryo Maturation:

Transformed embryogenic clusters were cultured for one-three weeks at 26 °C in SB196 under cool white fluorescent (Phillips cool white Econowatt F40/CW/RS/EW) and Agro (Phillips F40 Agro) bulbs (40 watt) on a 16:8 hrphotoperiod with light intensity of 90-120 $\mu E/m^2s$. After this time embryo clusters were removed to a solid agar media, SB166, for 1week. Then subcultured to medium SB103 for 3 weeks. Alternatively, embryo clusters were removed to SB228 (SHaM) liquid media, 35 mL in 250 mL Erlenmeyer flask, for 2-3 weeks. Tissue cultured in SB228 was maintained on a rotary shaker, 130 rpm, 26 °C with cool white fluorescent lights on 16:8 h day/night photoperiod at light intensity of 60-85 µE/m2/s. During this period, individual embryos were removed from the clusters and screened for alterations in their fatty acid compositions as described *supra*.

Media Recipes:

SB 196 - FN Lite Liquid Proliferation Medium (per liter)

MS FeEDTA - 100x Stock 1	10 mL
MS Sulfate - 100x Stock 2	10 mL
FN Lite Halides - 100x Stock 3	10 mL
FN Lite P, B, Mo - 100x Stock 4	10 mL
B5 vitamins (1 mL/L)	1.0 mL
2,4-D (10mg/L final concentration)	1.0 mL

KNO_3	2.83 gm
$(NH_4)_2SO_4$	0.463 gm
Asparagine	1.0 gm
Sucrose (1%)	10 gm
pH 5.8	

FN Lite Stock Solutions

Stock Number		1000 mL	500 mL
1	MS Fe EDTA 100x Stock		
	Na_2 EDTA*	3.724 g	1.862 g
	$FeSO_4 – 7H_2O$	2.784 g	1.392 g
*Add first, dissolve in dark bottle while stirring			
2	MS Sulfate 100x stock		
	$MgSO_4 - 7H_2O$	37.0 g	18.5 g
	$MnSO_4 - H_2O$	1.69 g	0.845 g
	$ZnSO_4 - 7H_2O$	0.86 g	0.43 g
	$CuSO_4 - 5H_2O$	0.0025 g	0.00125 g
3	FN Lite Halides 100x Stock		
	$CaCl_2 - 2H_2O$	30.0 g	15.0 g
	KI	0.083 g	0.0715 g
	$CoCl_2 - 6H_2O$	0.0025 g	0.00125 g
4	FN Lite P, B, Mo 100x Stock		
	KH_2PO_4	18.5 g	9.25 g
	H_3BO_3	0.62 g	0.31 g
	$Na_2MoO_4 - 2H_2O$	0.025 g	0.0125 g

SB1 Solid Medium (per liter)

1 package MS salts (Gibco/ BRL – Cat. No. 11117-066)

1 mL B5 vitamins 1000X stock

31.5 g Glucose

2 mL 2,4-D (20 mg/L final concentration)

pH 5.7

8 g TC agar

SB199 Solid Medium (per liter)

1 package MS salts (Gibco/ BRL – Cat. No. 11117-066)

1 mL B5 vitamins 1000X stock

30g Sucrose

4 ml 2,4-D (40 mg/L final concentration)

pH 7.0

2 gm Gelrite

SB 166 Solid Medium (per liter)

1 package MS salts (Gibco/ BRL – Cat. No. 11117-066)

1 mL B5 vitamins 1000X stock

60 g maltose

750 mg $MgCl_2$ hexahydrate

5 g Activated charcoal

pH 5.7

2 g Gelrite

SB 103 Solid Medium (per liter)

1 package MS salts (Gibco/ BRL – Cat. No. 11117-066)

1 mL B5 vitamins 1000X stock

60 g maltose

750 mg MgCl2 hexahydrate

pH 5.7

2 g Gelrite

SB 71-4 Solid Medium (per liter)

1 bottle Gamborg's B5 salts w/ sucrose (Gibco/ BRL – Cat. No. 21153-036)

pH 5.7

5 g TC agar

2,4-D Stock

Obtain premade from Phytotech Cat. No. D 295 – concentration 1 mg/mL

B5 Vitamins Stock (per 100 mL)

Store aliquots at -20 °C

10 g Myo-inositol

100 mg Nicotinic acid

100 mg Pyridoxine HCl

1 g Thiamine

If the solution does not dissolve quickly enough, apply a low level of heat via the hot stir plate.

SB 228- Soybean Histodifferentiation & Maturation (SHaM) (per liter)

DDI H2O	600ml
FN-Lite Macro Salts for SHaM 10X	100ml
MS Micro Salts 1000x	1ml
MS FeEDTA 100x	10ml
CaCl 100x	6.82ml
B5 Vitamins 1000x	1ml
L-Methionine	0.149g
Sucrose	30g
Sorbitol	30g

Adjust volume to 900 mL

pH 5.8

Autoclave

Add to cooled media (≤30C):

*Glutamine (Final conc. 30mM) 4%	110 mL

*Note: Final volume will be 1010 mL after glutamine addition.

Because glutamine degrades relatively rapidly, it may be preferable to add immediately prior to using media. Expiration 2 weeks after glutamine is added; base media can be kept longer w/o glutamine.

FN-lite Macro for SHAM 10X- Stock #1 (per liter)

$(NH_4)2SO_4$ (Ammonium Sulfate)	4.63g
KNO_3 (Potassium Nitrate)	28.3g
$MgSO_4*7H_20$ (Magnesium Sulfate Heptahydrate)	3.7g
KH_2PO_4 (Potassium Phosphate, Monobasic)	1.85g

Bring to volume

Autoclave

MS Micro 1000X- Stock #2 (per 1 liter)

H_3BO_3 (Boric Acid)	6.2g
$MnSO_4*H_2O$ (Manganese Sulfate Monohydrate)	16.9g
ZnSO4*7H20 (Zinc Sulfate Heptahydrate)	8.6g
Na_2MoO_4*2H20 (Sodium Molybdate Dihydrate)	0.25g
$CuSO_4*5H_20$ (Copper Sulfate Pentahydrate)	0.025g
$CoCl_2*6H_20$ (Cobalt Chloride Hexahydrate)	0.025g
KI (Potassium Iodide)	0.8300g

Bring to volume

Autoclave

FeEDTA 100X- Stock #3 (per liter)

Na_2EDTA* (Sodium EDTA)	3.73g
$FeSO_4*7H_20$ (Iron Sulfate Heptahydrate)	2.78g

*EDTA must be completely dissolved before adding iron.

Bring to Volume

Solution is photosensitive. Bottle(s) should be wrapped in foil to omit light.

Autoclave

Ca 100X- Stock #4 (per liter)

$CaCl_2*2H_20$ (Calcium Chloride Dihydrate)	44g

Bring to Volume

Autoclave

B5 Vitamin 1000X- Stock #5 (per liter)

Thiamine*HCl	10g
Nicotinic Acid	1g
Pyridoxine*HCl	1g
Myo-Inositol	100g
Bring to Volume	
Store frozen	

4% Glutamine- Stock #6 (per liter)

DDI water heated to 30°C	900ml
L-Glutamine	40g
Gradually add while stirring and applying low heat.	
Do not exceed 35°C.	
Bring to Volume	
Filter Sterilize	
Store frozen *	

*Note: Warm thawed stock in 31°C bath to fully dissolve crystals.

Oil analysis:

Oil content of somatic embryos was measured using NMR. Briefly lyophilized soybean somatic embryo tissue was pulverized in genogrinder vial as described previously (Example 2). 20 – 200 mg of tissue powder were transferred to NMR tubes. Oil content of the somatic embryo tissue powder was calculated from the NMR signal as described in Example 2. A total of 29 and 26 event were generated with plasmids pKS120 and pKS423, respectively and oil content of somatic embryos was measured (Table 18)

Table 18

Oil content of soybean somatic embryos generated with pKS120 or pKS423

event ID	plasmid	% oil (NMR)	event ID	plasmid	% oil (NMR)
2598-14	pKS120	7.3	2599-6	pKS423	8.0
2598-17	pKS120	6.1	2599-2	pKS423	6.5
2598-21	pKS120	6.0	2599-24	pKS423	6.3
2598-7	pKS120	5.6	2599-1	pKS423	6.2

2598-2	pKS120	5.2	2599-13	pKS423	6.1
2598-18	pKS120	4.7	2599-22	pKS423	6.0
2598-23	pKS120	4.7	2599-3	pKS423	5.7
2598-26	pKS120	4.6	2599-4	pKS423	5.7
2598-27	pKS120	4.5	2599-19	pKS423	5.5
2598-8	pKS120	4.4	2599-20	pKS423	5.4
2598-13	pKS120	4.3	2599-10	pKS423	5.4
2598-6	pKS120	4.3	2599-25	pKS423	5.3
2598-10	pKS120	4.3	2599-11	pKS423	5.2
2598-22	pKS120	4.2	2599-9	pKS423	5.2
2598-9	pKS120	4.1	2599-16	pKS423	5.2
2598-30	pKS120	4.0	2599-12	pKS423	4.8
2598-28	pKS120	3.9	2599-7	pKS423	4.4
2598-1	pKS120	3.5	2599-15	pKS423	4.2
2598-19	pKS120	3.1	2599-14	pKS423	4.0
2598-29	pKS120	3.1	2599-8	pKS423	3.8
2598-24	pKS120	2.9	2599-23	pKS423	3.7
2598-12	pKS120	2.8	2599-17	pKS423	3.5
2598-5	pKS120	2.7	2599-26	pKS423	3.3
2598-15	pKS120	2.5	2599-5	pKS423	3.2
2598-4	pKS120	2.4	2599-18	pKS423	3.1
2598-16	pKS120	2.4	2599-21	pKS423	3.0
2598-20	pKS120	2.4			
2598-11	pKS120	2.3			
2598-3	pKS120	2.2			
		average % oil			average % oil
		3.9			**4.9**

Table 18 demonstrates that total fatty acid content in soybean somatic embryos is increased as result of down-regulation of a soy HpaIL gene (Glyma09g21760).

EXAMPLE 19

Compositional analysis of arabidospis events transformed with DNA constructs for seed-preferred silencing of HpaIL genes

The example describes seed composition of transgenic events generated with pKR1482 HpaIL 3'UTR (SEQ ID NO: 27). It demonstrates that transformation with DNA constructs for silencing of genes encoding plastidic Hpa-like genes leads to increased oil content that is accompanied by a reduction in seed storage protein and soluble carbohydrate content.

Two transgenic events 14733 and 14734 were generated by agrobacterium-mediated transformation with pKR1482 HpaIL 3'UTR (SEQ ID NO:27) as described in Example 5.

T3 seed of K14733 and 14734 were germinated on selective plant growth media containing kanamycin. Kanamycin-resistant seedlings were transferred to soil and grown alongside untransformed control plants as described in Example 5. At maturity T4 seeds were bulk-harvested from transgenic lines and control plants and subjected to oil analysis by NMR as described in Example 2. The seed samples were subjected to compositional analysis of protein and soluble carbohydrate content of triplicate samples as described in Example 2. The results of this analysis are summarized in Table 19.

TABLE 19

Seed composition of arabidospis events transformed with DNA constructs for silencing of plastidic HpaIL genes

Genotype	Event ID	Oil (%, NMR)	Protein %	fructose (µg mg^{-1} seed)	glucose (µg mg^{-1} seed)
pKR1482 HpaIL 3'UTR (T4)	K14733	43.8	16.9	0.5	3.2
	WT	40.4	18.9	0.5	4.0
	Δ TG/WT %	8.4	-10.3	4.9	-20.5
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 HpaIL 3'UTR (T4)	K14733	15.0	0.4	1.6	21.4
	WT	15.3	0.5	1.8	22.7
	Δ TG/WT %	-2.0	-6.5	-8.8	-5.8
Genotype	Event ID	Oil (%, NMR)	Protein %	fructose (µg mg^{-1} seed)	glucose (µg mg^{-1} seed)
pKR1482 HpaIL 3'UTR (T4)	K14733	43.3	16.5	0.4	2.6
	WT	41.9	18.2	0.5	4.0
	Δ TG/WT %	3.3	-9.6	-10.5	-34.5

Genotype	Event ID	sucrose (µg mg^{-1} seed)	raffinose (µg mg^{-1} seed)	stachyose (µg mg^{-1} seed)	total soluble CHO (µg mg^{-1} seed)
pKR1482 HpaIL 3'UTR (T4)	K14733	15.0	0.4	1.4	20.6
	WT	14.9	0.4	1.6	22.0
	Δ TG/WT %	0.6	-3.2	-10.3	-6.5
Genotype	Event ID	Oil (%, NMR)	Protein %	fructose (µg mg^{-1} seed)	glucose (µg mg^{-1} seed)
pKR1482 HpaIL 3'UTR (T4)	K14734	41.4	17.0	0.5	4.4
	WT	38.0	20.6	0.4	4.7
	Δ TG/WT %	8.9	-17.4	7.8	-7.0
Genotype	Event ID	sucrose (µg mg^{-1} seed)	raffinose (µg mg^{-1} seed)	stachyose (µg mg^{-1} seed)	total soluble CHO (µg mg^{-1} seed)
pKR1482 HpaIL 3'UTR (T4)	K14734	14.5	0.4	1.7	22.2
	WT	15.3	0.5	1.5	23.0
	Δ TG/WT %	-5.0	-12.0	14.9	-3.7

Table 18 demonstrates that the oil increase associated with the presence of the pKR1482 HpaIL 3'UTR transgene (SEQ ID NO:27) 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 20

Compositional analysis of arabidospis events transformed with DNA constructs for seed-preferred over-expression of HpaIL genes

The example describes seed composition of transgenic events generated with pKR1478- At4g10750 (SEQ ID NO:51). It demonstrates that transformation with DNA constructs for seed-preferred overexpression genes encoding plastidic HpaI-

like genes leads to decreased oil content that is accompanied by increased seed storage protein content.

Primers HpaILORF FWD (SEQ ID NO:11) and HpaIL ORF REV (SEQ ID NO:12) were used to amplify the At4g10750 ORF from genomic DNA of Arabidopsis plants of the Landsberg *erecta genotype*. The PCR product was cloned into pENTR (Invitrogen, USA) to give pENTR- At4g10750 (SEQ ID NO:50). The HpaIL 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. A gel-purified DNA fragment of 2222 bp was excised from pENTR- At4g10750 with the restriction enzymes EcoR V and ApaL I and used in the recombination reaction. The sequence of the resulting plasmid pKR1478- At4g10750 is set forth as SEQ ID NO:51. The HpaIL ORF present in pKR1478- At4g10750 and its deduced amino acid sequence are set forth in SEQ ID NO: 52 and SEQ ID NO: 53, respectively. They represent the At4g10750 gene sequence of *Arabidopsis thaliana* of gentotype Landsberg *erecta*. Said sequences are 99.6 and 99.4 % identical to the nucleotide sequence and deduced amino acid sequence of SEQ ID NOs: 48 and 49, respectively. The latter represent the nucleotide and deduced amino acid sequence of the At4g10750 sequence of *Arabidopsis thaliana* of gentotype Columbia. As stated in the current example genomic DNA of *Arabidopsis thaliana* of genotype Landsberg *erecta and not* of genotype Columbia (as was stated erroneously in Example 4 of the instant specification) was used as a template to PCR amplify the ORF present in pENTR-At4g10750 and pKR1478- At4g10750.

pKR1478- At4g10750 (SEQ ID NO:51) was introduced into *Agrobacterium tumefaciens* NTL4 (Luo et al, *Molecular Plant-Microbe Interactions* (2001) 14(1):98-103) by electroporation. Briefly, 100 ng 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, 1kΩ and 25 µF. Cells were transferred to 1 mL LB medium and incubated for 3 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 250 mL 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- HpaIL 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 250 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. T1 Plants were grown to maturity alongside wt control plants and T2 seeds were harvested and oil content was analyzed by NMR as described above (Example 2).

TABLE 20

Seed oil content of T1 plants generated with binary vector pKR1478- At4g10750 for seed preferred overexpression of At4g10750

Construct	BARCODE	% oil	oil content % of WT	avg. oil content % of WT
pKR1478- At4g10750	K50660	40.6	99.9	
pKR1478- At4g10750	K50655	40.5	99.6	
pKR1478- At4g10750	K50672	40.4	99.5	
pKR1478- At4g10750	K50663	40.1	98.8	
pKR1478- At4g10750	K50661	39.0	96.0	

pKR1478- At4g10750	K50653	38.6	94.9	
pKR1478- At4g10750	K50669	38.4	94.5	
pKR1478- At4g10750	K50662	38.3	94.3	
pKR1478- At4g10750	K50667	38.0	93.6	
pKR1478- At4g10750	K50652	37.7	92.8	
pKR1478- At4g10750	**K50668**	31.5	77.5	94.7
wt	K50679	42.6		
wt	K50676	40.6		
wt	K50680	40.4		
wt	K50678	39.8		
wt	K50677	39.6		
Construct	**BARCODE**	**% oil**	**oil content % of WT**	**avg. oil content % of WT**
pKR1478- At4g10750	K50696	41.2	100.4	
pKR1478- At4g10750	K50687	40.8	99.5	
pKR1478- At4g10750	K50697	40.3	98.1	
pKR1478- At4g10750	K50683	40.2	97.9	
pKR1478- At4g10750	K50686	40.1	97.8	
pKR1478- At4g10750	K50692	39.9	97.1	
pKR1478- At4g10750	K50681	39.5	96.3	
pKR1478- At4g10750	K50693	39.5	96.2	
pKR1478- At4g10750	**K50689**	38.2	93.2	
pKR1478- At4g10750	K50684	38.1	92.8	96.9
wt	K50701	42.3		
wt	K50698	41.1		
wt	K50700	40.7		
wt	K50699	40.1		

TABLE 21

Seed oil content of T2 plants generated with binary vector pKR1478- At4g10750 for seed preferred overexpression of At4g10750

event	**BARCODE**	**% oil**	**oil content % of WT**	**avg. oil content % of WT**
K50668	K53363	40.1	98.9	
K50668	K53361	40.0	98.7	
K50668	K53359	39.9	98.4	
K50668	K53375	39.1	96.4	
K50668	K53377	39.1	96.3	
K50668	K53376	38.8	95.6	

K50668	K53365	38.7	95.4	
K50668	K53362	37.7	92.9	
K50668	K53371	36.8	90.7	
K50668	K53370	36.6	90.2	
K50668	K53366	36.5	90.1	
K50668	K53369	36.0	88.8	
K50668	K53373	35.9	88.6	
K50668	K53372	31.9	78.7	92.8
wt	K53379	42.3		
wt	K53382	42.2		
wt	K53388	41.8		
wt	K53385	41.3		
wt	K53381	40.3		
wt	K53387	39.8		
wt	K53384	39.8		
wt	K53383	39.6		
wt	K53380	39.5		
wt	K53386	39.0		
Construct	**BARCODE**	**% oil**	**oil content % of WT**	**avg. oil content % of WT**
K50689	K53259	39.5	96.4	
K50689	K53257	39.0	95.2	
K50689	K53251	38.7	94.4	
K50689	K53253	38.3	93.5	
K50689	K53249	38.1	93.1	
K50689	K53256	37.7	92.1	
K50689	K53247	37.5	91.6	
K50689	K53248	37.2	90.7	
K50689	K53261	36.8	90.0	
K50689	K53260	36.8	89.8	
K50689	K53246	35.8	87.3	
K50689	K53258	35.4	86.5	91.7
wt	K53270	42.8		
wt	K53266	42.2		
wt	K53271	41.9		
wt	K53267	41.7		
wt	K53272	41.3		
wt	K53275	41.3		
wt	K53273	41.2		
wt	K53264	40.7		

wt	K53269	40.5		
wt	K53268	40.4		
wt	K53265	40.1		
wt	K53274	37.2		

T3 seed of events K50668 and K50689 with oil contents shown in Table 21 were combined and seed composition of the events was analyzed and compared to seed composition of untransformed WT plants gown in the same flat. Analysis of seed composition was performed as described in Example 2.

TABLE 22

Seed composition of arabidospis events transformed with DNA constructs for seed-preferred overexpression of plastidic HpaIL genes

Genotype	Event ID	Oil (%, NMR)	Protein %	fructose (µg mg^{-1} seed)	glucose (µg mg^{-1} seed)
pKR1478-At4g10750	K50668	35.9	20.6	0.5	5.7
	WT	39.3	18.5	0.6	3.9
	Δ TG/WT %	-8.7	11.6	-6.3	48.0
Genotype	Event ID	sucrose (µg mg^{-1} seed)	raffinose (µg mg^{-1} seed)	stachyose (µg mg^{-1} seed)	total soluble CHO (µg mg^{-1} seed)
pKR1478-At4g10750	K50668	15.0	0.5	1.5	24.0
	WT	17.2	0.5	1.5	24.2
	Δ TG/WT %	-12.4	4.1	0.0	-0.7
Genotype	Event ID	Oil (%, NMR)	Protein %	fructose (µg mg^{-1} seed)	glucose (µg mg^{-1} seed)
pKR1478-At4g10750	K50689	37.3	18.2	0.5	4.5
	WT	39.5	17.0	0.5	4.1
	Δ TG/WT %	-5.5	7.0	-2.1	10.1
Genotype	Event ID	sucrose (µg mg^{-1} seed)	raffinose (µg mg^{-1} seed)	stachyose (µg mg^{-1} seed)	total soluble CHO (µg mg^{-1} seed)

pKR1478-At4g10750	K50689	16.6	0.5	1.6	24.3
	WT	16.9	0.4	1.6	24.3
	Δ TG/WT %	-1.8	1.4	1.9	0.3

Tables 20 and 21 demonstrate that seed-preferred over-expression of HpaIL genes such as At4g10750 leads to a heritable reduction in seed oil content. Table 22 shows that this oil reduction is accompanied by an increase in seed storage protein.

EXAMPLE 21

Characterization of arabidospis events transformed with DNA constructs that contain the complete HpaIL gene ORF for seed-preferred silencing of HpaIL genes

The example describes seed composition of transgenic events generated with pKR1482 At4g10750 (SEQ ID NO:54). It demonstrates that transformation with DNA constructs for silencing of genes encoding plastidic Hpal-like genes such as At4g10750 leads to increased oil content.

5 μg of plasmid DNA of pENTR- At4g10750 (SEQ ID NO:50) was digested with PvuII. A restriction fragment of 1715 bp (derived from pENTR- At4g10750) was excised from an agarose gel. The entire protein coding sequence of the HpaIL gene represented by SEQ ID NO: 52 was inserted in the sense and anti-sense orientation into vector pKR1482 (SEQ ID NO:15) using LR clonase (Invitrogen) according to the manufacturer's instructions, to give pKR1482- At4g10750 (SEQ ID NO:54). Transgenic arabidopsis lines were generated as described previously (Example 19) and oil content of transgenic T2 seed and untransformed control seed from plants grown in the same flat alongside the transgenic lines was analyzed by NMR as described in Example 2.

TABLE 23

Seed oil content of T1 plants generated with binary vector pKR1482- At4g10750 for seed preferred silencing of At4g10750

Construct	**BARCODE**	**% oil**	**oil content % of WT**	**avg. oil content % of WT**
pKR1482- At4g10750	**K50819**	43.4	106.5	
pKR1482- At4g10750	K50816	43.3	106.2	

pKR1482- At4g10750	K50807	43.2	106.0	
pKR1482- At4g10750	K50814	43.2	106.0	
pKR1482- At4g10750	K50796	42.9	105.3	
pKR1482- At4g10750	K50817	42.9	105.2	
pKR1482- At4g10750	K50798	42.8	105.0	
pKR1482- At4g10750	K50808	42.7	104.7	
pKR1482- At4g10750	K50800	42.6	104.4	
pKR1482- At4g10750	K50804	42.1	103.3	
pKR1482- At4g10750	K50820	42.0	103.0	
pKR1482- At4g10750	K50794	41.9	102.8	
pKR1482- At4g10750	K50810	41.7	102.4	
pKR1482- At4g10750	K50818	41.6	102.1	
pKR1482- At4g10750	K50815	41.6	101.9	
pKR1482- At4g10750	K50801	41.4	101.5	
pKR1482- At4g10750	K50806	41.0	100.7	
pKR1482- At4g10750	K50799	40.8	100.0	
pKR1482- At4g10750	K50821	40.8	100.0	
pKR1482- At4g10750	K50812	40.8	100.0	
pKR1482- At4g10750	K50795	40.6	99.6	
pKR1482- At4g10750	K50813	40.1	98.5	
pKR1482- At4g10750	K50809	39.5	96.9	
pKR1482- At4g10750	K50802	39.1	95.9	
pKR1482- At4g10750	K50811	38.9	95.4	
pKR1482- At4g10750	K50803	38.7	94.9	
pKR1482- At4g10750	K50805	37.8	92.8	
pKR1482- At4g10750	K50822	35.7	87.5	
pKR1482- At4g10750	K50797	35.1	86.2	100.5
wt	K50824	41.7		
wt	K50823	41.5		
wt	K50825	41.1		
wt	K50826	40.8		
wt	K50827	38.6		

T2 seed of event K50819 were germinated on selective plant growth media containing kanamycin, planted in soil alongside WT plants and grown to maturity. T3 seed oil content was measured by NMR.

TABLE 24

Seed oil content of T2 plants generated with binary vector pKR1482-At4g10750 for seed preferred silencing of At4g10750

Event ID	BARCODE	% oil	oil content % of WT	avg. oil content % of WT
K50819	**K53078**	42.2	108.4	
K50819	**K53070**	41.9	107.8	
K50819	**K53064**	41.8	107.5	
K50819	**K53069**	41.5	106.8	
K50819	K53077	40.9	105.2	
K50819	K53062	40.4	104.0	
K50819	K53075	40.3	103.8	
K50819	K53068	39.4	101.4	
K50819	K53063	39.2	100.8	
K50819	K53061	39.1	100.6	
K50819	K53066	39.0	100.4	
K50819	K53060	38.8	99.9	
K50819	K53076	38.6	99.2	
K50819	K53072	38.4	98.8	
K50819	K53067	38.3	98.5	
K50819	K53065	38.1	97.9	
K50819	K53074	38.0	97.7	
K50819	K53073	37.6	96.7	
K50819	K53071	37.0	95.1	101.6
wt	K53084	40.1		
wt	K53083	39.9		
wt	K53087	39.8		
wt	K53089	39.6		
wt	K53080	39.6		
wt	K53079	39.5		
wt	K53086	39.1		
wt	K53081	39.0		
wt	K53090	38.8		
wt	K53088	38.3		
wt	K53082	38.2		
wt	K53091	37.9		
wt	K53085	35.6		

T3 seed of lines K53078, K53070, K53064 and K53069 derived event from event K50819 were germinated on selective plant growth media containing kanamycin, planted in soil alongside WT plants and grown to maturity. T4 Seed oil content was measured by NMR.

5

TABLE 25

Seed oil content of T3 plants generated with binary vector pKR1482- At4g10750 for seed preferred silencing of At4g10750

T4 line ID	BARCODE	% oil	oil content % of WT	avg. oil content % of WT
K50819/K53078	K59771	44.8	115.6	
K50819/K53078	K59778	43.7	112.9	
K50819/K53078	K59780	43.4	112.2	
K50819/K53078	K59786	42.9	110.9	
K50819/K53078	K59775	42.5	109.8	
K50819/K53078	K59774	42.5	109.7	
K50819/K53078	K59777	42.5	109.7	
K50819/K53078	K59769	42.4	109.5	
K50819/K53078	K59784	42.2	109.1	
K50819/K53078	K59781	42.2	108.9	
K50819/K53078	K59770	42.0	108.4	
K50819/K53078	K59785	41.6	107.6	
K50819/K53078	K59779	41.2	106.3	
K50819/K53078	K59768	40.9	105.7	
K50819/K53078	K59776	40.6	104.9	
K50819/K53078	K59765	40.1	103.5	
K50819/K53078	K59772	39.9	103.0	
K50819/K53078	K59767	38.8	100.2	
K50819/K53078	K59773	38.7	100.1	
K50819/K53078	K59783	37.9	98.0	
K50819/K53078	K59766	37.3	96.3	
K50819/K53078	K59782	36.8	95.0	106.2
wt	K59791	40.6		
wt	K59792	39.9		
wt	K59788	39.7		
wt	K59787	38.9		
wt	K59789	38.7		

wt	K59793	38.4		
wt	K59794	38.3		
wt	K59790	35.3		
T4 line ID	**BARCODE**	**% oil**	**oil content % of WT**	**avg. oil content % of WT**
K50819/K53070	K58902	45.4	112.4	
K50819/K53070	K58906	45.1	111.7	
K50819/K53070	K58886	45.0	111.6	
K50819/K53070	K58896	44.9	111.4	
K50819/K53070	K58894	44.8	111.0	
K50819/K53070	K58904	44.6	110.6	
K50819/K53070	K58895	44.4	110.1	
K50819/K53070	K58887	44.3	109.8	
K50819/K53070	K58889	44.1	109.3	
K50819/K53070	K58888	44.1	109.3	
K50819/K53070	K58897	43.9	108.7	
K50819/K53070	K58901	43.9	108.7	
K50819/K53070	K58905	43.7	108.2	
K50819/K53070	K58903	43.5	107.8	
K50819/K53070	K58891	43.5	107.7	
K50819/K53070	K58900	43.2	107.0	
K50819/K53070	K58892	42.8	106.0	
K50819/K53070	K58893	42.8	105.9	
K50819/K53070	K58898	42.1	104.3	
K50819/K53070	K58899	41.7	103.3	
K50819/K53070	K58890	38.7	95.9	
K50819/K53070	K58907	35.2	87.1	107.2
wt	K58915	42.3		
wt	K58908	42.1		
wt	K58909	42.0		
wt	K58913	41.5		
wt	K58911	41.2		
wt	K58910	40.2		
wt	K58914	40.0		
wt	K58917	39.7		
wt	K58916	38.2		
wt	K58912	36.4		

T4 line ID	BARCODE	% oil	oil content % of WT	avg. oil content % of WT
K50819/K53064	K58936	45.2	114.3	
K50819/K53064	K58934	45.2	114.3	
K50819/K53064	K58933	45.1	114.0	
K50819/K53064	K58919	45.0	113.7	
K50819/K53064	K58921	44.9	113.6	
K50819/K53064	K58924	44.8	113.2	
K50819/K53064	K58927	44.5	112.5	
K50819/K53064	K58925	44.4	112.2	
K50819/K53064	K58918	44.1	111.4	
K50819/K53064	K58928	44.0	111.3	
K50819/K53064	K58929	43.7	110.3	
K50819/K53064	K58926	43.0	108.8	
K50819/K53064	K58930	43.0	108.6	
K50819/K53064	K58923	42.7	108.0	
K50819/K53064	K58931	42.6	107.7	
K50819/K53064	K58922	41.4	104.5	
K50819/K53064	K58920	39.0	98.5	
K50819/K53064	K58935	38.8	98.0	
K50819/K53064	K58932	37.2	94.1	108.9
wt	K58945	41.8		
wt	K58944	40.8		
wt	K58941	40.5		
wt	K58937	40.5		
wt	K58939	40.2		
wt	K58942	39.5		
wt	K58943	38.7		
wt	K58940	38.2		
wt	K58938	35.8		
T4 line ID	**BARCODE**	**% oil**	**oil content % of WT**	**avg. oil content % of WT**
K50819/K53069	K60122	44.1	112.0	
K50819/K53069	K60127	44.1	112.0	
K50819/K53069	K60125	44.0	111.6	
K50819/K53069	K60123	43.8	111.3	
K50819/K53069	K60117	43.7	111.1	
K50819/K53069	K60120	43.6	110.7	
K50819/K53069	K60114	43.4	110.4	

K50819/K53069	K60113	43.4	110.2	
K50819/K53069	K60128	43.3	110.1	
K50819/K53069	K60124	43.3	110.1	
K50819/K53069	K60126	43.1	109.6	
K50819/K53069	K60119	42.4	107.7	
K50819/K53069	K60115	42.2	107.2	
K50819/K53069	K60116	41.6	105.6	
K50819/K53069	K60121	41.0	104.0	
K50819/K53069	K60129	39.4	100.2	
K50819/K53069	K60118	39.3	99.9	108.4
wt	K60138	41.1		
wt	K60135	40.5		
wt	K60132	40.2		
wt	K60131	39.8		
wt	K60130	39.6		
wt	K60134	39.6		
wt	K60133	39.6		
wt	K60137	39.2		
wt	K60136	34.7		

Table 23-25 show that silencing of HpaI-like genes such as At4g10750 using hairpin constructs comprised of the entire protein coding region of the gene lead to a heritable oil increase. InT4 lines that are homozygous for the T-DNA insertion the average oil content was 6-9 % higher than that of WT control plants.

EXAMPLE 22

Combination HpaL-like gene silencing and over-expression of acylCoA: diacylglycerol acyltransferase (DGAT) in soybean somatic embryos

The example describes generation of transgenic soybean somatic embryos that contain either constructs for down-regulation of soybean HpaIL genes, or constructs for overexpression of acylCoA: diacylglycerol acyltransferase (DGAT) genes. Moreover transgenic soybean somatic embryos are described that show both, over-expression of DGAT and down-regulation of soybean HpaIL genes. The latter embryos have a total oil content that exceeds the oil content observed in embryos harboring single transgenes responsible for DGAT overexpression or silencing of HpaIL genes.

Patent application number WO 2009143398 A1 describes plasmid KS387 (SEQ ID NO:55) for co-expression of DGAT1 and DGAT2 genes of *Yarrowia*

lipolytica in developing soybean seed. Plasmids KS120 (SEQ ID NO:46) and KS423 (SEQ ID NO:45) are described in Example 18.

For experiments MSE 2650 and MSE 2653 purified plasmid DNA of KS120 and KS423 was used for generation of transgenic soybean somatic embryos exactly as described in Example 18. For experiment MSE2651 a 10:1 ratio of DNA of plasmids KS387 and KS120 was used for generation of transgenic soybean somatic embryos exactly as described in Example 18. For experiment MSE2652 a 10:1 ratio of DNA of plasmids KS387 and KS423 was used for generation of transgenic soybean somatic embryos exactly as described in Example 18. Oil content of lyophilized soybean somatic embryos of experiments MSE2650-2653 was analyzed by NMR as described in Examples 2 and is reported in Table 26.

TABLE 26

Oil content of somatic embryos generated with plasmids KS120, KS387, KS423 or a combination thereof

experiment name	plasmid	event id	% oil	average % oil
MSE 2650	KS120	K52130	6.6	
		K52140	5.6	
		K52137	5.5	
		K52136	5.2	
		K52131	4.8	
		K52139	4.8	
		K52122	4.7	
		K52146	4.6	
		K52133	4.5	
		K52151	4.4	
		K52145	4.3	
		K52147	4.0	
		K52132	3.8	
		K52143	3.8	
		K52124	3.7	
		K52149	3.6	
		K52129	3.6	
		K52134	3.6	
		K52128	3.6	
		K52150	3.3	
		K52148	3.3	
		K52138	3.2	
		K52127	3.2	

		K52135	3.1	
		K52126	3.1	
		K52123	3.0	
		K52142	3.0	
		K52141	2.9	
		K52121	2.9	
		K52144	2.8	
		K52125	2.8	**3.9**
experiment name	plasmid	event id	% oil	average % oil
2653	KS423	K52237	9.8	
		K52243	7.8	
		K52214	6.2	
		K52227	6.1	
		K52233	5.9	
		K52236	5.6	
		K52231	5.4	
		K52228	5.3	
		K52230	5.0	
		K52238	4.9	
		K52239	4.8	
		K52218	4.7	
		K52215	4.7	
		K52220	4.7	
		K52235	4.6	
		K52242	4.5	
		K52232	4.5	
		K52221	4.5	
		K52241	4.4	
		K52229	4.1	
		K52226	4.0	
		K52224	3.8	
		K52217	3.8	
		K52225	3.5	
		K52222	3.4	
		K52240	3.3	
		K52219	3.0	
		K52219	3.0	
		K52234	2.6	
		K52234	2.6	
		K52223	2.5	
		K52223	2.5	
		K52216	2.2	**4.5**

experiment name	plasmid	event id	% oil	average % oil
2651	KS387/KS120	K52171	9.7	
		K52166	9.5	
		K52159	9.0	
		K52179	8.9	
		K52178	8.7	
		K52158	8.0	
		K52153	8.0	
		K52157	7.9	
		K52173	7.7	
		K52155	7.5	
		K52163	7.3	
		K52180	7.3	
		K52177	7.2	
		K52182	6.5	
		K52170	5.6	
		K52169	5.5	
		K52165	5.4	
		K52161	5.3	
		K52172	5.1	
		K52175	5.0	
		K52174	4.4	
		K52181	4.2	
		K52168	3.7	
		K52152	3.6	
		K52156	3.3	
		K52164	3.3	
		K52176	3.0	
		K52160	2.7	**6.2**

experiment name	plasmid	event id	% oil	average % oil
2652	KS387/KS423	K52188	12.7	
		K52190	11.5	
		K52186	11.1	
		K52206	11.1	
		K52197	10.6	
		K52203	10.4	
		K52193	9.5	
		K52183	9.5	
		K52211	9.1	
		K52201	9.0	
		K52204	8.5	
		K52187	8.3	
		K52213	7.7	
		K52207	6.7	

K52196	6.7	
K52208	6.5	
K52194	6.3	
K52192	6.2	
K52185	6.2	
K52202	6.1	
K52212	5.4	
K52198	5.0	
K52191	4.9	
K52195	4.8	
K52199	4.3	
K52189	4.1	
K52205	3.9	
K52200	3.8	
K52210	3.6	
K52184	3.2	
K52209	3.0	**7.1**

In summary Table 26 demonstrates that transformation with constructs for silencing of HpaIL genes increased average oil content of soybean somatic embryos by 14.5 %, transformation with constructs for co-expression of yarrowia DGAT genes increased average oil of soybean somatic embryos by 58 % and transformation with constructs for co-expression of yarrowia DGAT genes **and** silencing of HpaIL genes increased oil content by 81%. Thus the additive effect of both metabolic engineering approaches on soybean oil content provides clear evidence that HpaL gene silencing and DGAT overexpression direct carbohydrates towards oil biosynthesis through independent, i.e. distinct routes.

EXAMPLE 24

Seed-preferred silencing of HpaIL genes in soybean using artificial miRNAs

The example describes the construction of a plasmid vector for soybean transformation. The plasmid provides seed-preferred expression of two artificial microRNAs that both target soybean gene Glyma09g21760 (SEQ ID NO: 30). Soybean somatic embryos transformed with plasmid constructs containing either one of the two artificial microRNA showed increased oil content compared to embryos that harbor a control plasmid.

Vectors were made to silence HpaIL genes genes using an artificial microRNA largely as described in US patent application no.:12,335,717, filed December 16, 2008. The following briefly explains the procedure.

Design of Artificial MicroRNA sequences

Artificial microRNAs (amiRNAs) that would have the ability to silence the desired target genes were designed largely according to rules described in Schwab R, et al. (2005) *Dev Cell* 8: 517-27. To summarize, microRNA sequences are 21 nucleotides in length, start at their 5'-end with a "U", display 5' instability relative to their star sequence which is achieved by including a C or G at position 19, and their 10th nucleotide is either an "A" or an "U". An additional requirement for artificial microRNA design was that the amiRNA have a high free delta-G as calculated using the ZipFold algorithm (Markham, N. R. & Zuker, M. (2005) *Nucleic Acids Res.* 33: W577-W581.) The DNA sequence corresponding to the first amiRNA (ALDO A) that was used to silence aldolase is set forth in SEQ ID NO:56. The DNA sequence corresponding to the second amiRNA (ALDO B) that was used to silence the aldolase gene is set forth in SEQ ID NO:57.

Design of an artificial star sequences

"Star sequences" are those that base pair with the amiRNA sequences, in the precursor RNA, to form imperfect stem structures. To form a perfect stem structure the star sequence would be the exact reverse complement of the amiRNA. The soybean precursor sequence as described in "Novel and nodulation-regulated microRNAs in soybean roots" Subramanian S, Fu Y, Sunkar R, Barbazuk WB, Zhu JK, Yu O BMC Genomics. 9:160(2008) and accessed on mirBase (Conservation and divergence of microRNA families in plants" Dezulian T, Palatnik JF, Huson DH, Weigel D (2005) Genome Biology 6:P13) was folded using mfold (M. Zuker (2003) *Nucleic Acids Res.* 31: 3406-15; and D.H. Mathews, J. et al. (1999) *J. Mol. Biol.* 288: 911-940). The miRNA sequence was then replaced with the amiRNA sequence and the endogenous star sequence was replaced with the exact reverse complement of the amiRNA. Changes in the artificial star sequence were introduced so that the structure of the stem would remain the same as the endogenous structure. The altered sequence was then folded with mfold and the original and altered structures were compared by eye. If necessary, further alternations to the artificial star sequence were introduced to maintain the original structure. The first amiRNA star sequence (ALDO A star) that was used to silence aldolase is set forth as SEQ ID NO:58. The 2nd amiRNA star sequence (ALDO B star) that was used to silence aldolase is set forth as SEQ ID NO:59.

Conversion of Genomic MicroRNA Precursors to Artificial MicroRNA Precursors

Genomic miRNA precursor genes were converted to amiRNA precursors using In-Fusion™ as described above. In brief, the microRNA 159 precursor (SEQ ID NO:60) was altered to include Pme I sites immediately flanking the star and microRNA sequences to form the in-fusion ready microRNA 159 precursor (SEQ ID NO:61). This sequence was cloned into the Not I site of KS126 to form the in-fusion ready microRNA 159-KS126 plasmid (SEQ ID NO:62). KS126 is described in PCT Publication No. WO 04/071467.

The microRNA 159 precursor (SEQ ID NO:60) was used as a PCR template. The primers (gmir159ALDO A1, SEQ ID NO:63 and gmir159ALDO A2, SEQ ID NO:64) were designed according to the protocol provided by Clontech and do not leave any footprint of the Pme I sites after the In-Fusion recombination reaction. The sequence of resulting amplified 159-ALDO A DNA is shown in SEQ ID NO:65.

The microRNA 159 precursor (SEQ ID NO:60) was used as a PCR template. The primers (gmir159ALDO B1, SEQ ID NO:66 and gmir159ALDO B2, SEQ ID NO:67) were designed according to the protocol provided by Clontech and do not leave any footprint of the Pme I sites after the In-Fusion recombination reaction. The sequence of resulting amplified 159-ALDO B DNAs is shown in SEQ ID NO:68.

The sequences of 159-ALDO A (SEQ ID NO:65) and 159-ALDO B (SEQ ID NO:67) were recombined into the in-fusion ready microRNA 159-KS126 plasmid (SEQ ID NO:62) digested with PmeI. This was done using protocols provided with the In-Fusion™ kit. The resulting plasmid are 159 ALDO A-KS126 (SEQ ID NO:69) and 159 ALDO B-KS126 (SEQ ID NO:70).

Plasmid DNA of 159 ALDO A-KS126 (SEQ ID NO:69) 159 ALDO B-KS126 (SEQ ID NO:70) and a control plasmid KS120 (SEQ ID NO:46) was used for transformation soybean cell suspensions and subsequent generation of soybean somatic embryos as described in Example 18. Oil content of soybean somatic embryos was measured by NMR and is summarized in Table 27.

TABLE 27

Oil content of somatic embryos generated with plasmids KS120, 159 ALDO A-KS126 and 159 ALDO B-KS126

experiment name	plasmid	event id	% oil	average % oil
2672	KS120	K54409	5.7	
		K54408	4.6	
		K54386	4.5	
		K54389	4.4	
		K54401	4.2	
		K54405	4.2	
		K54393	4.1	
		K54394	3.9	
		K54387	3.7	
		K54406	3.6	
		K54390	3.6	
		K54400	3.6	
		K54398	3.5	
		K54397	3.5	
		K54410	3.5	
		K54395	3.5	
		K54382	3.4	
		K54381	3.4	
		K54391	3.3	
		K54399	3.2	
		K54385	3.2	
		K54402	3.0	
		K54407	3.0	
		K54388	3.0	
		K54392	3.0	
		K54396	2.8	
		K54404	2.8	**3.6**
experiment name	plasmid	event id	% oil	average % oil
2670	159 ALDO A-KS126	K54326	11.4	
		K54346	9.5	
		K54322	8.6	
		K54325	8.2	
		K54340	7.1	
		K54345	6.8	
		K54324	6.8	
		K54329	6.4	
		K54333	6.4	
		K54337	6.2	
		K54343	6.0	
		K54336	6.0	
		K54331	5.9	

		K54349	5.9	
		K54328	5.9	
		K54341	5.8	
		K54327	5.8	
		K54339	5.6	
		K54348	5.3	
		K54334	5.2	
		K54332	5.0	
		K54342	4.8	
		K54323	4.7	
		K54338	4.3	
		K54347	4.3	
		K54320	4.2	
		K54335	4.0	
		K54321	3.7	
		K54330	3.4	
		K54344	3.2	**5.9**
experiment name	plasmid	event id	% oil	average % oil
2671	159 ALDO B-KS126	K54378	8.0	
		K54368	7.6	
		K54350	7.0	
		K54363	5.6	
		K54361	5.5	
		K54373	5.5	
		K54360	5.4	
		K54371	5.3	
		K54375	5.2	
		K54365	5.0	
		K54357	4.9	
		K54355	4.4	
		K54372	4.4	
		K54358	4.3	
		K54380	4.2	
		K54352	4.2	
		K54374	4.2	
		K54376	4.2	
		K54364	4.0	
		K54369	3.8	
		K54356	3.7	
		K54379	3.7	
		K54362	3.7	
		K54353	3.6	
		K54370	3.6	
		K54359	3.5	

K54354	3.4	
K54366	3.3	
K54377	3.3	
K54351	3.2	
K54367	2.7	**4.5**

Table 27 demonstrates that total fatty acid content in soybean somatic embryos was increased as result of down-regulation of a soy HpaIL gene (Glyma09g21760) mediated by expression of artificial microRNAs targeting said gene.

EXAMPLE 25

Expression of bacterial HpaI and plant HpaIL genes in E.coli and analysis of enzyme activity of recombinantly-produced proteins

The example describes expression of bacterial HpaI and plant HpaIL genes in E. coli, purification of recombinantly-produced bacterial HpaI and plant HpaIL enzymes and analysis of enzyme properties such as divalent ion and pH requirements and kinetic properties with pyruvate and acetaldehyde substrates. The example demonstrates that plant HpaIL enzymes, like distantly related bacterial HpaI enzymes can catalyze aldol additions using pyruvate and short chain aldehydes. Similar to bacterial HpaI enzymes, catalysis by plant HpaI-like enzymes requires presence of divalent ions. In these reactions catalytic efficiency (Kcat/Km) of plant HpaI-like enzyme is about 20-30 fold lower than of bacterial HpaI enzymes. Finally it is shown that in plant HpaI-like enzymes, similar to prokaryotic HpaI enzyme, a certain n-terminal arginine residue is required for aldol addition enzyme activity.

The amino acid sequence of the arabidopsis HpaIL protein derived from At4g10750 SEQ ID NO:48 was analyzed using ChloroP at the online ChloroP 1.1. Server, which predicts the presence of chloroplast transit peptides (cTP) in protein sequences and the location of potential cTP cleavage sites. An n-terminal chloroplast targeting signal peptide of 65 amino acids was identified. Primers AthHpaIL fwd (SEQ ID NO:71) and AthHpaIL rev (SEQ ID NO:72) and plasmid DNA of pKR1478- At4g10750 (SEQ ID NO:51) were used to PCR amplify a fragment of the At4g10750 transcript that corresponds to the processed, plastid-localized

At4g10750 gene product. PCR products were cloned into pGEM®-T Easy to give pGEM®-T At4g10750 (SEQ ID NO:73).

pGEM®-T At4g10750 was digested with NcoI SalI. A restriction fragment of 890 bp was gel-purified and ligated to NcoI SalI linearized plasmid DNA of pET28a (Novagen/EMD4Biosciences, NJ, USA) to give pET28a At4g10750 (SEQ ID NO:74). The amino acid sequence of the At4g10750 gene product including a c-terminal pET28A-derived hexa-histidine tag are set forth as SEQ ID NO:75.

The amino acid sequence of soy HpaIL protein derived from Glyma09g21760 was analyzed using ChloroP (*supra*). An n-terminal chloroplast targeting signal peptide of 60 amino acids was identified. Primers Soy HpaIL fwd (SEQ ID NO:76) and Soy HpaIL rev (SEQ ID NO:77) and plasmid of applicants EST clone sfp1n.pk022.m19 were used to PCR amplify a fragment of the Glyma09g21760 transcript that corresponds to the processed, plastid-localized Glyma09g21760 gene product. PCR products were cloned into pGEM®-T Easy to give pGEM®-T Glyma09g21760 (SEQ ID NO:78).

pGEM®-T Glyma09g21760 was digested with NdeI SacI. A restriction fragment of 884 bp was gel-purified and ligated to NdeI SacI-linearized plasmid DNA of pET29a (Novagen/EMD4Biosciences, NJ, USA) to give pET29a Glyma09g21760 (SEQ ID NO:79). The amino acid sequence of the Glyma09g21760 gene product including a c-terminal pET29a-derived hexa-histidine tag are set forth as SEQ ID NO:80 .

The amino acid sequence of the rice HpaIL protein derived from Os09g36030 was analyzed using ChloroP (*supra*). An n-terminal chloroplast targeting signal peptide of 47 amino acids was identified. Primers Rice HpaIL fwd (SEQ ID NO:81) and Rice HpaIL rev (SEQ ID NO:82) and plasmid applicants EST clone rdi2c.pk005.c17 were used to PCR amplify a fragment of the Os09g36030 transcript that corresponds to the processed, plastid-localized Os09g36030 gene product. PCR products were cloned into pGEM®-T Easy to give pGEM®-T Os09g36030 (SEQ ID NO:83).

pGEM®-T Os09g36030 was digested with NcoI and HindIII. A restriction fragment of 884 bp was gel-purified and ligated to NcoI HindIII-linearized plasmid DNA of pET28a (supra) to give pET28a Os09g36030 (SEQ ID NO:84). The amino

acid sequence of the Os09g36030 gene product including a c-terminal pET28a-derived hexa-histidine tag are set forth as SEQ ID NO:85 .

A bacterial HpaI gene was amplified from genomic DNA of *Pseudomonas putida* strain DSM 12585 described in: Muheim, A.; Lerch, K. Towards a high-yield bioconversion of ferulic acid to vanillin. Applied Microbiology and Biotechnology (1999), 51(4), 456-461. Briefly, a PCR product of 900 bp was PCR amplified with primer PP FWD (SEQ ID NO:86) and PP REV (SEQ ID NO:87) and genomic DNA of *Pseudomonas putida* strain DSM 12585. PCR products were cloned into pCR Blunt-II-Topo (Invitrogen, USA) according to manufacturer instructions to give pCR blunt HpaI PP (SEQ ID NO:88). Recombinant plasmid DNA was sequenced. The DNA and deduced amino acid sequence sequence of HpaI ORF of *Pseudomonas putida* strain DSM 12585 henceforth named HpaI PP is set forth as SEQ ID NO:89 and SEQ ID NO:90, respectively. A DNA fragment for expression cloning into pET29a was generated. Briefly, a PCR product of 800 bp was PCR amplified with primer HpaI PP FWD (SEQ ID NO:91) and HpaI PP REV (SEQ ID NO:92) and plasmid DNA of pCR blunt HpaI PP (SEQ ID NO:88). PCR products were cloned into pGEM®-T Easy to give pGEM®-T HpaI PP (SEQ ID NO:93).

pGEM®-T HpaI PP (SEQ ID NO:93) was digested with NdeI SacI. A restriction fragment of 800 bp was gel-purified and ligated to NdeI SacI-linearized plasmid DNA of pET29a (supra) to give pET29a HpaI PP (SEQ ID NO:94). The amino acid sequence of the HpaI PP gene product including a c-terminal pET29a-derived hexa-histidine tag are set forth as SEQ ID NO:95.

Competent E. coli cells of strain Rosetta™(DE3)pLysS (Novagen/EMD4Biosciences, NJ, USA) were transformed with pET28a At4g10750 (SEQ ID NO:74) using electroporation. Four 500 mL flasks each containing 250 mL of LB medium supplemented with 50 μg/mL kanamycin were inoculated with E coli cells of strain Rosetta™(DE3)pLysS carrying pET28a At4g10750 (SEQ ID NO:74). The culture was grown at 37 °C until a cell density ($OD_{\lambda=600nm}$) of 0.6 was achieved. The cultures were cooled to 16 °C on ice. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was then added to a final concentration of 0.1 mM followed by continued culture at 16 °C for 36 h. Cells were harvested by centrifugation (5000xg, 10 min) and resupended in 30 mL of 50 mM Hepes/KOH (pH 8), 0.5M NaCl, 10 mM Imidazole, 2mM DTT. The cell suspension was passed twice through a French

press and cleared by centrifugation (30000xg, 20 min, 4°C). The enzyme extract (30 mL) was buffer-exchanged in 2.5 mL aliquots on PD10 columns (GE Healthcare, USA) into 50 mM Hepes/KOH (pH 8), 500 mM NaCl, 20 mM Imidazole. Buffer-exchanged extract (40 mL) was loaded onto a HiTrap chelating HP column with 5 mL gel bed volume (GE Healthcare, Uppsala, Sweden). The HiTrap chelating HP column had previously been charged with Ni^{2+} according to manufacturer instructions. The column was developed at a flow rate of 2 mL/min at 22°C as follows: Solvent A (50 mM Hepes/KOH (pH 8), 500mM NaCl, 20 mM Imidazole), Solvent B (50 mM Hepes/KOH (pH 8), 500mM NaCl, 500mM Imidazole); 0-20 min 0% B, 20-35 min 20% B, 35-50 min (linear gradient) 20-100% B, 50-55min 100% B, 55-60 min 0% B. 1.5 mL fractions were collected from beginning to end of the linear imidazole gradient. 10 μL fractions were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (PAGE). A protein of 33 kDA was observed in fractions 11- 15 indicating that the expected 6xhistagged At4g10750 protein variant (SEQ ID NO:75) was present in the *E. coli* extract and could be purified by Ni^{2+} affinity column chromatography.

Bacterial HpaI enzymes catalyze aldol addition reactions using substrates such as pyruvate and acetaldehyde leading to the formation of 4-hydroxy-2-oxovalerate with said substrates (Wang, Weijun; Baker, Perrin; Seah, Stephen Y. K. Comparison of Two Metal-Dependent Pyruvate Aldolases Related by Convergent Evolution: Substrate Specificity, Kinetic Mechanism, and Substrate Channeling. Biochemistry (2010), 49(17), 3774-3782).

Next, Ni^{2+} affinity column chromatography fractions of protein extracts of *E coli* expressing pET28a At4g10750 were assayed for this activity. Briefly 20 μL of fraction 9 with undetectable levels of 6xhistagged At4g10750 protein and fraction 12 with very high levels of said protein were combined with 80 mM pyruvate, 80 mM acetaldehyde, 100mM Hepes, KOH, pH 8, 2mM $CoCl_2$ in a final volume of 100 μL. Reactions were incubated at 27 °C for 20 min. 20 μL of HCl was added, samples were incubated at 100 °C for 3 min , quenched on ice, cleared by centrifugation an analyzed by HPLC as follows.

10 μL of sample were separated using an HP Agilent 1100 HPLC system equipped with an Aminex fast-acid analysis ion-exchange column, 100mm length, 7.8mm diameter (Biorad, Hercules, CA, USA). The column was developed at a flow

rate of 2 mL min^{-1} using 10 mM H_2SO_4 and reaction products were detected using a diode array UV detector at λ= 210nm and λ= 230nm.

Reaction products of fraction 9 contained two predominant peaks with retention times of 2.1 and 2.6 min. The former shared its retention time with an unmodified pyruvate standard. The latter absorbed more strongly at λ= 230nm. Applicants assume that the latter compound is the lactone of γ-hydroxy γ-methyl α-keto glutarate. It is well established that two molecules of pyruvate can spontaneously react at alkaline pH to form one molecule of γ-hydroxy γ-methyl α-keto glutarate (Formation of γ -hydroxy-γ -methylglutamic acid from a common impurity in pyruvic acid. Goldfine, H Biochimica et Biophysica Acta (1960), 40:557-9). Compared to reaction products obtained with fraction 9, products obtained with fraction 12 showed a great reduction in pyruvate and putative γ-hydroxy γ-methyl α-keto glutarate-lactone peaks and a new, predominant peak with a retention time of 6.06 min that absorbed strongly at λ= 230nm. Subsequent enzyme assays revealed that the production of the compound with a retention time of 6.06 min was only observed when both pyruvate and acetaldehyde were provided to recombinantly produced At4g10750 protein. Applicants conclude that, most likely, the peak with a retention time of 6.06 min is the lactonized product of an aldol addition reaction between acetaldehyde and pyruvate catalyzed by the At4g10750 protein. The expected molecule would be the lactone of 4-hydroxy-2-oxovalerate with an expected MW of 114.1. Large scale synthesis of the compound with a retention time of 6.06 min was performed as follows. The final reaction mixture contained 80 mM pyruvate, 80 mM acetaldehyde, 100 mM Hepes/KOH pH 8, 2mM $CoCl_2$, and 100 μL of fraction 12, corresponding to approximately 50 μg of recombinantly-produced At4g10750 protein in a final volume of 5 mL. The reaction was incubated at 27 ^{o}C for 16h. One mL of concentrated HCl was added and the reaction mixture was heated to 100 ^{o}C for 3 min. The aqueous reaction was extracted three times with 5 mL of ethylacetate. 100 μL of the ethylacetate extract was dried down using N_2 and resuspended in 100 μL of water. Ten μL were analyzed by HPLC as described above. When analyzed by HPLC the ethyl acetate extracted reaction products contained one predominant peak with a retention time of 6.06 min. The entire ethylactetate extract was dried down and analyzed by

GC/electron impact MS. Reaction products were dissolved in 3mL of dichloromethane and 1 μL was subjected to GC/MS analysis. Reaction products were separated on a DB-5MS column using a Agilent 6890 GC using the following temperatue program: initial temperature 70 °C, hold 4 min, temperature ramp 10 °C min^{-1} to 300 °C, hold 7 min. Mass spectroscopy was performed using an Hewlett-Packard mass selective detector according to manufacturer instructions with ms source and ms quad temperatures at -150 °C and 230 °C, respectively. Total icon current chromatograms revealed the presence of two peaks with retention times of 6.5 and 6.9 min. Mass spectra of booth peaks reveal the presence of a molecular ion with a mass to charge ratio (m/z) of 114.1. MS spectra of both peaks contain fragments with m/z of 86.2, 58.1 and 43.1. 6.5 and 6.8 min peaks show differences in the mass spectra the former contains two fragments with m/z of 69.2 and 99.0, whereas the latter instead contains two fragments with m/z of 71.2 and 97.2. A second preparative synthesis of the compound with a retention time of 6.06 min was performed in a volume of 5 mL exactly as described above. Reaction products were resuspended in D_2O and analyzed by ^1H NMR (500MHz). The following shifts were identified : chem. shift ^1H = 5.25 (=C-H), 2.07 ($-CH_2-$),1.60 (-CH), 1.50 (C-H), 1.23 ($-CH_3$). In summary, both GC/MS and ^1H NMR analysis of reaction products generated with recombinantly-produced At4g10750 enzyme, pyruvate and acetaldehyde reveal that the reaction product with a HPLC retention time of 6.06 min is a mixture of 5-Methyl-dihydro-furan-2,3-dione and the enol form of this molecule which is 3-Hydroxy-5-methyl-5H-furan-2-one. These molecules can also be referred to as 4-hydroxy-2-oxovalerate-lactone and the enol isomer of said molecule.

Purification of recombinantly-produced At4g10750 protein

Four 500 mL flasks each containing 333 mL of LB medium supplemented with 50 μg/mL kanamycin were inoculated with E coli cells of strain Rosetta™(DE3)pLysS carrying pET28a At4g10750 (SEQ ID NO:74). The culture was grown at 37 °C until a cell density ($OD_{\lambda=600nm}$) of 0.6 was achieved. The cultures were cooled to 16 °C on ice. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was then added to a final concentration of 0.1 mM followed by continued culture at 16 °C for 36 h. Cells were harvested by centrifugation (5000xg, 10 min) and

recombinantly-produced At4g10750 protein was purified from the cell pellet as described above. 10 μL aliquots of Ni^{2+} affinity chromatography fractions were analyzed by SDS page and coomassie staining. Fractions containing the recombinantly produced At4g10750 protein were pooled and buffer exchanged into 20 mM Hepes/NaOH, pH8, 5% w/v glycerol and stored at -80 °C. The protein concentration of the buffer exchanged Ni^{2+} affinity chromatography fractions was measured at 3.5 mg mL^{-1} using the Bradford assay (Biorad, USA) according to manufacturer instructions. Visual inspection of overloaded Coomassie-stained gels indicated that the purified recombinant At4g10750 protein was at least 95% pure. In summary approximately 36 mg of recombinantly produced At4g10750 protein were purified from 750 mg of total protein of E coli cells carrying pET28a At4g10750.

Purification of recombinantly-produced Glyma09g21760 protein

Four 500 mL flasks each containing 333 mL of LB medium supplemented with 50 μg/mL kanamycin were inoculated with E coli cells of strain Rosetta™(DE3)pLysS carrying pET29a Glyma09g21760 (SEQ ID NO:79). The culture was grown at 37 °C until a cell density ($OD_{\lambda=600nm}$) of 0.6 was achieved. The cultures were cooled to 16 °C on ice. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was then added to a final concentration of 0.1 mM followed by continued culture at 16 °C for 36 h. Cells were harvested by centrifugation (5000xg, 10 min) and recombinantly produced Glyma09g21760 protein was purified from the cell pellet as described above. 10 uL aliquots of Ni^{2+} affinity chromatography fractions were analyzed by SDS page and coomassie staining. Fractions containing the recombinantly-produced Glyma09g21760 protein were pooled and buffer exchanged into 20 mM Hepes/NaOH, pH8, 5% w/v glycerol and stored at -80 °C. The protein concentration of the buffer exchanged Ni^{2+} affinity chromatography fractions was measured at 1.96 mg mL^{-1} using the Bradford assay (Biorad, USA) according to manufacturer instructions. Visual inspection of overloaded Coomassie-stained gels indicated that the purified recombinant Glyma09g21760 protein was at least 95% pure. In summary approximately 20 mg of recombinantly produced Glyma09g21760 protein were purified from 750 mg of total protein of E coli cells carrying pET29a Glyma09g21760.

Purification of recombinantly produced Os09g36030 protein

Six 500 mL flasks each containing 333 mL of LB medium supplemented with 50 μg/mL kanamycin were inoculated with E coli cells of strain Rosetta™(DE3)pLysS carrying pET28a Os09g36030 (SEQ ID NO:84). The culture was grown at 37 °C until a cell density ($OD_{\lambda=600nm}$) of 0.6 was achieved. The cultures were cooled to 16 °C on ice. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was then added to a final concentration of 0.1 mM followed by continued culture at 16 °C for 36 h. Cells were harvested by centrifugation (5000xg, 10 min) and recombinantly-produced Os09g36030 protein was purified from the cell pellet as described above. 10 μL aliquots of Ni^{2+} affinity chromatography fractions were analyzed by SDS page and coomassie staining. Fractions containing the recombinantly-produced Os09g36030 protein were pooled and buffer exchanged into 20 mM Hepes/NaOH, pH8, 5 % w/v glycerol and stored at -80 °C. The buffer exchanged Ni^{2+} affinity chromatography fractions were further concentrated to a final volume of 1.3 mL using Centriprep YM 10 centrifugal concentrators (Millipore, USA) according to manufacturer instructions. The protein concentration of the buffer exchanged Ni^{2+} affinity chromatography fractions was measured at 1.9 mg mL^{-1} using the Bradford assay (Biorad, USA) according to manufacturer instructions. Visual inspection of overloaded Coomassie-stained gels indicated that the purified recombinant Os09g36030 protein was at least 50% pure. In summary, approximately 2.5 mg of recombinantly-produced Os09g36030 protein were purified from 900 mg of total protein of *E coli* cells carrying pET28a Os09g36030.

Purification of recombinantly produced *P. putida* HpaI protein

Five 500 mL flasks each containing 333 mL of LB medium supplemented with 50 μg/mL kanamycin were inoculated with E coli cells of strain Rosetta™(DE3)pLysS carrying pET29a HpaI PP (SEQ ID NO:94). The culture was grown at 37 °C until a cell density ($OD_{\lambda=600nm}$) of 0.6 was achieved. The cultures were cooled to 16 °C on ice. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was then added to a final concentration of 0.1 mM followed by continued culture at 16 °C for 36 h. Cells were harvested by centrifugation (5000xg, 10 min) and recombinantly-produced *P. putida* HpaI protein was purified from the cell pellet as described above. 10 μL aliquots of Ni^{2+} affinity chromatography fractions were analyzed by

SDS page and coomassie staining. Fractions containing the recombinantly-produced *P. putida* HpaI protein were pooled and buffer exchanged into 20 mM Hepes/NaOH, pH8, 5% w/v glycerol and stored at -80 °C. The protein concentration of the buffer exchanged Ni^{2+} affinity chromatography fractions was measured at 6.75 mg mL^{-1} using the Bradford assay (Biorad, USA) according to manufacturer instructions. Visual inspection of overloaded Coomassie-stained gels indicated that the purified recombinant *P. putida* HpaI protein was at least 95% pure. In summary approximately 110 mg of recombinantly-produced P. putida HpaI protein were purified from 1200 mg of total protein of *E. coli* cells carrying pET29a HpaI PP.

HPLC-based quantitation of the lactone of 4-hydroxy-2-oxovalerate

100 μL samples (50mM Hepes/KOH, pH8) with pyruvate concentrations of 2.5, 5, 7.5, 10, 15 and 20 mM were supplemented with 20 μL of concentrated HCl, heated to 100°C for 3 min, quenched in ice water. 30 μL of each sample was separated by HPLC on a Fast Acid ion exchange column as described above. The peak areas of pyruvate (RT 2.1 min, l=210nm) and putatively identified γ-hydroxy γ–methyl α-keto glutarate-lactone (RT 2.1 min, l=230nm) were recorded. The sum of both peak areas is henceforth referred to as Peak area Pyr-E. The same range of pyruvate concentrations (2.5-20 mM) was incubated with 2.5 μg of recombinantly-produced *P. putida* HpaI protein in a final volume of 100 μL in the presence of 50mM Hepes/KOH, pH8, 20 mM acetaldehyde, 2mM $CoCl_2$ for 15min at 27 °C. Enzyme reactions were supplemented with 20 μL of concentrated HCl, heated to 100 °C for 3 min, quenched in ice water. 30 μL of each sample was separated by HPLC on a Fast Acid ion exchange column as described above. Peak areas of pyruvate (RT 2.1 min, λ=210nm), putatively identified γ-hydroxy γ–methyl α-keto glutarate-lactone (RT 2.1 min, λ =230nm) and the lactone of 4-hydroxy-2-oxovalerate (RT 6.1 min, λ =230nm) were recorded. The sum of peak areas of pyruvate (RT 2.1 min, λ =210nm) and putatively identified γ-hydroxy γ–methyl α-keto glutarate-lactone (RT 2.1 min, λ =230nm) derived from the enzyme treated sample is henceforth referred to as Peak area Pyr+E. The concentration of pyruvate consumed by the HpaI enzyme-catalyzed aldol condensation of pyruvate and acetaldehyde in each enzyme-treated sample can be calculated using the following

formula: [Pyr consumed]=(Peak area Pyr-E - Peak area Pyr+E)/Peak area Pyr-E x [Pyr T_0]. In each HpaI-enzyme-treated sample [Pyr consumed] calculated in this manner is equal to the concentration of 4-hydroxy-2-oxovalerate-lactone produced.

TABLE 28

Values for HPLC-based quantitation of 4-hydroxy-2-oxovalerate-lactone

μM pyruvate [Pyr T_0]	μM pyruvate consumed [Pyr consumed]	peak area 4-hydroxy-2-oxovalerate-lactone (mAU, λ=230nm)
2500	392	1260
5000	549	2069
7500	799	2981
10000	1220	3774
15000	1279	4582
20000	1751	5562

A calibration curve for quantitation of 4-hydroxy-2-oxovalerate-lactone was established using the values shown in column 2 and 3 of Table 28. According to this calibration curve the concentration (μM) of 4-hydroxy-2-oxovalerate-lactone in a given HPLC sample can be calculated by multiplying the peak area of RT 6.1 at λ=230 with 0.2993.

Divalent ion requirements of a bacterial HpaI enzyme and plant-derived HpaIL enzymes

Divalent ion requirements of recombinantly-produced At4g10750 enzyme were determined as follows: 25 μg of recombinantly-produced At4g10750 protein were incubated in the presence of no added divalent ion or 2mM of either $CoCl_2$, $CaCl_2$, $MnCl_2$ or $MgCl_2$ in a final volume of 100 μL of 10mM pyruvate, 10mM acetaldehyde. 50 mM Hepes/KOH, pH 8.0. Enzyme assays were performed at 27 °C for 20 min. Reactions were stopped by addition of 20 μL concentrated HCl and incubation at 100 °C for 3 min. Reaction products were separated by HPLC and 4-hydroxy-2-oxovalerate-lactone production was quantitated using the previously described calibration curve.

TABLE 29

Divalent ion requirements of recombinantly-produced At4g10750 enzyme

ion	specific activity (pmol s^{-1} mg^{-1} protein)
no ion	152.5
Co^{2+}	2866.6
Ca^{2+}	647.0
Mn^{2+}	2832.3
Mg^{2+}	5489.6

Divalent ion requirements of recombinantly-produced Glyma09g21760 enzyme were determined as follows: 50 μg of recombinantly-produced Glyma09g21760 protein were incubated in the presence of no added divalent ion or 2mM of either $CoCl_2$, $CaCl_2$, $MnCl_2$ or $MgCl_2$ in a final volume of 100 μL of 10mM pyruvate, 10mM acetaldehyde. 50 mM Hepes/KOH, pH 8.0. Enzyme assays were performed at 27 °C for 20 min. Reactions were stopped by addition of 20 μL concentrated HCl and incubation at 100 °C for 3 min. Reaction products were separated by HPLC and 4-hydroxy-2-oxovalerate-lactone production was quantitated using the previously described calibration curve.

TABLE 30

Divalent ion requirements of recombinantly produced Glyma09g21760 enzyme

ion	specific activity (pmol s^{-1} mg^{-1} protein)
no ion	0.0
Co^{2}+	986.1
Ca^{2}+	847.1
Mn^{2+}	1155.8
Mg^{2+}	1890.2

Divalent ion requirements of recombinantly-produced Os09g36030 enzyme were determined as follows: 25 μg of recombinantly-produced Os09g36030 protein were incubated in the presence of no added divalent ion or 2mM of either $CoCl_2$, $CaCl_2$, $MnCl_2$ or $MgCl_2$ in a final volume of 100 μL of 10mM pyruvate, 10mM acetaldehyde and 50 mM Hepes/KOH, pH 8.0. Enzyme assays were performed at 27 °C for 18 min. Reactions were stopped by addition of 20 μL concentrated HCl and incubation at 100 °C for 3 min. Reaction products were separated by HPLC and

4-hydroxy-2-oxovalerate-lactone production was quantitated using the previously described calibration curve.

TABLE 31

Divalent ion requirements of recombinantly produced Os09g36030 enzyme

ion	specific activity (pmol s^{-1} mg^{-1} protein)
no ion	89.3
Co^{2}+	1344.4
Ca^{2}+	119.7
Mn^{2+}	1669.1
Mg^{2+}	667.6

Divalent ion requirements of recombinantly-produced *P. putida* HpaI enzyme

2.5 μg of recombinantly-produced *P. putida* HpaI protein were incubated in the presence of no added divalent ion or 2mM of either $CoCl_2$, $CaCl_2$, $MnCl_2$ or $MgCl_2$ in a final volume of 100 μL of 10mM pyruvate, 10mM acetaldehyde and 50 mM Hepes/KOH, pH 7.25. Enzyme assays were performed at 27 °C for 15 min. Reactions were stopped by addition of 20 μL concentrated HCl and incubation at 100 °C for 3 min. Reaction products were separated by HPLC and 4-hydroxy-2-oxovalerate-lactone production was quantitated using the previously described calibration curve.

TABLE 32

Divalent ion requirements of recombinantly-produced *P. putida* HpaI enzyme

ion	specific activity (pmol s^{-1} mg^{-1} protein)
no ion	5155.7
Co^{2}+	95304.8
Ca^{2}+	6282.4
Mn^{2+}	62343.5
Mg^{2+}	38026.4

pH requirements of a bacterial HpaI enzyme and plant-derived HpaIL enzymes

pH requirements of recombinantly produced At4g10750 enzyme were determined as follows: 25 μg of recombinantly produced At4g10750 protein were incubated in the presence 100 mM Bis-Tris-Propane/HCl covering a pH range form 7-9.5 in 0.25 pH point increments in a final volume of 100 μL of 10mM pyruvate,

10mM acetaldehyde and 2mM $MgCl_2$. Enzyme assays were performed at 27 ^oC for 10 min. Reactions were stopped by addition of 20 μL concentrated HCl and incubation at 100 ^oC for 3 min. Reaction products were separated by HPLC and 4-hydroxy-2-oxovalerate-lactone production was quantitated using the previously described calibration curve.

TABLE 33

pH requirements of recombinantly produced At4g10750 enzyme

pH	specific activity (pmol s^{-1} mg^{-1} protein)
7	955.3
7.25	1121.4
7.5	1206.1
7.75	1354.3
8	1380.5
8.25	1375.8
8.75	1254.8
9.25	1129.3
9.5	1077.2

pH requirements of recombinantly produced Glyma09g21760 enzyme were determined as follows: 25 μg of recombinantly produced Glyma09g21760protein were incubated in the presence 100 mM Bis-Tris-Propane/HCl covering a pH range form 7-9.5 in 0.25 pH point increments in a final volume of 100 μL of 10mM pyruvate, 10mM acetaldehyde and 2mM $MgCl_2$. Enzyme assays were performed at 27 ^oC for 10 min. Reactions were stopped by addition of 20 μL concentrated HCl and incubation at 100 ^oC for 3 min. Reaction products were separated by HPLC and 4-hydroxy-2-oxovalerate-lactone production was quantitated using the previously described calibration curve.

TABLE 34

pH requirements of recombinantly produced Glyma09g21760 enzyme

pH	specific activity (pmol s^{-1} mg^{-1} protein)
7	883.1
7.25	1151.0
7.5	1442.3
7.75	2161.3
8	2321.3
8.25	2273.5

8.5	2046.1
8.75	1894.0
9	1882.3
9.25	1813.0
9.5	1798.5

pH requirements of recombinantly produced Os09g36030 enzyme were determined as follows: 25 μg of recombinantly produced Os09g36030 protein were incubated in the presence 100 mM Bis-Tris-Propane/HCl covering a pH range form 7-9.5 in 0.25 pH point increments in a final volume of 100 μL of 10mM pyruvate, 10mM acetaldehyde and 2mM $MnCl_2$. Enzyme assays were performed at 27 °C for 10 min. Reactions were stopped by addition of 20 μL concentrated HCl and incubation at 100 °C for 3 min. Reaction products were separated by HPLC and 4-hydroxy-2-oxovalerate-lactone production was quantitated using the previously described calibration curve.

TABLE 35

pH requirements of recombinantly produced Os09g36030 enzyme

pH	specific activity (pmol s^{-1} mg^{-1} protein)
7	271.6
7.25	265.9
7.5	337.7
7.75	406.9
8	461.8
8.25	486.4
8.5	456.2
8.75	408.7
9	382.4
9.25	251.4
9.5	144.8

pH requirements of recombinantly produced *P. putida* HpaI enzyme were determined as follows: 2.5 μg of recombinantly produced *P. putida* HpaI protein were incubated in the presence 100 mM MES/KOH or Bis-Tris-Propane/HCl covering a pH range of 4.5-8 in a final volume of 100 μL of 10mM pyruvate, 10mM acetaldehyde and 2mM $CoCl_2$. Enzyme assays were performed at 27 °C for 15 min. Reactions were stopped by addition of 20 μL concentrated HCl and incubation at 100 °C for 3 min. Reaction products were separated by HPLC and 4-hydroxy-2-

oxovalerate-lactone production was quantitated using the previously described calibration curve.

TABLE 36

pH requirements of recombinantly produced *P. putida* HpaI enzyme

pH	Buffer	specific activity (pmol s^{-1} mg^{-1} protein)
4.5	MES	31770.7
5	MES	39220.9
5.5	MES	48823.0
6	MES	46708.5
6.5	MES	54927.3
7	HEPES	75939.2
7.25	HEPES	78898.3
7.5	HEPES	73605.4
7.75	HEPES	68430.3
8	HEPES	65166.7

Analysis of kinetic properties of recombinantly produced At4g10750 enzyme with pyruvate

Kinetic properties of recombinantly produced At4g10750 enzyme with the substrate pyruvate were determined as follows. Formation of 4-hydroxy-2-oxovalerate was assayed using 25 μg of recombinantly produced At4g10750 enzyme in a final volume of 100 microliters in the presence of 50 mM Hepes/KOH pH8, 10 mM acetaldehyde, 2 mM $MgCl_2$ and pyruvate concentrations ranging from 1 to 32 mM. Assays were performed for 15 min at 27 ^{o}C. 4-hydroxy-2-oxovalerate was quantitated by HPLC analysis as described above. Under these conditions, apparent Km and Vmax values were 3.79 mM and 3139 pmol s^{-1} mg^{-1} protein, respectively. These parameters were determined using the Eadie–Hofstee plot by plotting velocity/substrate concentration versus velocity using velocities determined at pyruvate concentrations of 2, 4, 6, 8 and 10 mM. In this plot an estimate of the Km is provided as the slope of the line representing the linear regression curve through the points and the Vmax by the intercept of the regression curve with the y axis.

Analysis of kinetic properties of recombinantly produced At4g10750 enzyme with acetaldehyde

Kinetic properties of recombinantly produced At4g10750 enzyme with the substrate acetaldehyde were determined as follows. Formation of 4-hydroxy-2-oxovalerate was assayed using 25 μg of recombinantly produced At4g10750 enzyme in a final volume of 100 μL in the presence of 50 mM Hepes/KOH pH8, 10 mM pyruvate, 2 mM $MgCl_2$ and acetaldehyde concentrations ranging from 1 to 32 mM. Assays were performed for 15 min at 27 °C. 4-hydroxy-2-oxovalerate was quantitated by HPLC analysis as described above. Under these conditions, apparent Km and Vmax values were 1.37 mM and 2253 pmol s^{-1} mg^{-1} protein, respectively. These parameters were determined using the Eadie–Hofstee plot by plotting velocity/substrate concentration versus velocity using velocities determined at pyruvate concentrations 1,2,4,6,8,10,12,14,16,18 and 32 mM. In this plot an estimate of the Km is provided as the slope of the line representing the linear regression curve through the points and the Vmax by the intercept of the regression curve with the y axis.

Analysis of kinetic properties of recombinantly produced Glyma09g21760 enzyme with pyruvate

Kinetic properties of recombinantly produced Glyma09g21760 enzyme with the substrate pyruvate were determined as follows. Formation of 4-hydroxy-2-oxovalerate was assayed using 25 μg of recombinantly produced At4g10750 enzyme in a final volume of 100 μL in the presence of 50 mM Hepes/KOH pH8, 10 mM acetaldehyde, 2 mM $MgCl_2$ and pyruvate concentrations ranging from 1 to 32 mM. Assays were performed for 15 min at 27 °C. 4-hydroxy-2-oxovalerate was quantitated by HPLC analysis as described above. Under these conditions, apparent Km and Vmax values were 13.2 mM and 7853 pmol s^{-1} mg^{-1} protein, respectively. These parameters were determined using the Hofstee plot by plotting velocity/substrate concentration versus velocity using velocities determined at pyruvate concentrations of 2,4,6,8,10,12,14,16,18 and 32. In this plot an estimate of the Km is provided as the slope of the line representing the linear regression curve through the points and the Vmax by the intercept of the regression curve with the y axis.

Analysis of kinetic properties of recombinantly produced Glyma09g21760 enzyme with acetaldehyde

Kinetic properties of recombinantly produced Glyma09g21760 enzyme with the substrate pyruvate were determined as follows. Formation of 4-hydroxy-2-oxovalerate was assayed using 25 μg of recombinantly produced Glyma09g21760 enzyme in a final volume of 100 μL in the presence of 50 mM Hepes/KOH pH8, 10 mM pyruvate, 2 mM $MgCl_2$ and acteladehyde concentrations ranging from 1 to 32 mM. Assays were performed for 15 min at 27 °C. 4-hydroxy-2-oxovalerate was quantitated by HPLC analysis as described above. Under these conditions, apparent Km and Vmax values were 1.74 mM and 5366 pmol s^{-1} mg^{-1} protein, respectively. These parameters were determined using the Hofstee plot by plotting velocity/substrate concentration versus velocity using velocities determined at pyruvate concentrations 2,4,6,8,10,12,14,16,18 and 32 mM. In this plot an estimate of the Km is provided as the slope of the line representing the linear regression curve through the points and the Vmax by the intercept of the regression curve with the y axis.

Analysis of kinetic properties of recombinantly produced Os09g36030 enzyme with pyruvate

Kinetic properties of recombinantly produced Os09g36030 enzyme with the substrate pyruvate were determined as follows. Formation of 4-hydroxy-2-oxovalerate was assayed using 25 μg of recombinantly produced Os09g36030 enzyme in a final volume of 100 μL in the presence of 50 mM Hepes/KOH pH8, 10 mM acetaldehyde, 2 mM $MnCl_2$ and pyruvate concentrations ranging from 1 to 32 mM. Assays were performed for 15 min at 27 °C. 4-hydroxy-2-oxovalerate was quantitated by HPLC analysis as described above. Under these conditions, apparent Km and Vmax values were 7.5 mM and 2104 pmol s^{-1} mg^{-1} protein, respectively. These parameters were determined using the Hofstee plot by plotting velocity/substrate concentration versus velocity using velocities determined at pyruvate concentrations of 1,2,4,6,8,10,12,14,16,18 and 32. In this plot an estimate of the Km is provided as the slope of the line representing the linear regression curve through the points and the Vmax by the intercept of the regression curve with the y axis.

Analysis of kinetic properties of recombinantly produced Os09g36030 enzyme with acetaldehyde

Kinetic properties of recombinantly produced Os09g36030 enzyme with the substrate pyruvate were determined as follows. Formation of 4-hydroxy-2-oxovalerate was assayed using 25 μg of recombinantly produced Os09g36030 enzyme in a final volume of 100 μL in the presence of 50 mM Hepes/KOH, pH 7.25, 10 mM pyruvate, 2 mM $MgCl_2$ and acteladehyde concentrations ranging from 1 to 32 mM. Assays were performed for 15 min at 27 °C. 4-hydroxy-2-oxovalerate was quantitated by HPLC analysis as described above. Under these conditions, apparent Km and Vmax values were 1.48 mM and 1304 pmol s^{-1} mg^{-1} protein, respectively. These parameters were determined using the Hofstee plot by plotting velocity/substrate concentration versus velocity using velocities determined at pyruvate concentrations 1,2,4,6,8 and 10 mM. In this plot an estimate of the Km is provided as the slope of the line representing the linear regression curve through the points and the Vmax by the intercept of the regression curve with the y axis.

Analysis of kinetic properties of recombinantly produced *P. putida* HpaI enzyme with pyruvate

Kinetic properties of recombinantly produced *P. putida* HpaI enzyme with the substrate pyruvate were determined as follows. Formation of 4-hydroxy-2-oxovalerate was assayed using 1.25 μg of recombinantly-produced *P. putida* HpaI enzyme in a final volume of 100 μL in the presence of 50 mM Hepes/KOH, pH 7.25, 10 mM acetaldehyde, 2 mM $CoCl_2$ and pyruvate concentrations ranging from 1 to 80 mM. Assays were performed for 15 min at 27 °C. 4-hydroxy-2-oxovalerate was quantitated by HPLC analysis as described above. Under these conditions, apparent Km and Vmax values were 5.3 mM and 122866 pmol s^{-1} mg^{-1} protein, respectively. These parameters were determined using the Hofstee plot by plotting velocity/substrate concentration versus velocity using velocities determined at pyruvate concentrations of 1, 2, 4, 6, 8, 10, 12, 16, 20, 25, 30, 40, 60 and 80 mM. In this plot an estimate of the Km is provided as the slope of the line representing the linear regression curve through the points and the Vmax by the intercept of the regression curve with the y axis.

Analysis of kinetic properties of recombinantly-produced *P. putida* HpaI enzyme with acetaldehyde

Kinetic properties of recombinantly produced *P. putida* HpaI enzyme with the substrate pyruvate were determined as follows. Formation of 4-hydroxy-2-oxovalerate was assayed using 1.25 μg of recombinantly produced *P. putida* HpaI enzyme in a final volume of 100 microliters in the presence of 50 mM Hepes/KOH, pH 7.25, 10 mM pyruvate, 2 mM $CoCl_2$ and acteladehyde concentrations ranging from 1 to 80 mM. Assays were performed for 15 min at 27 °C. 4-hydroxy-2-oxovalerate was quantitated by HPLC analysis as described above. Under these conditions, apparent Km and Vmax values were 2.98 mM and 93010 pmol s^{-1} mg^{-1} protein, respectively. These parameters were determined using the Eadie–Hofstee plot by plotting velocity/substrate concentration versus velocity using velocities determined at pyruvate concentrations of 1, 2, 4, 6, 8,10, 12, 16, 20, 25, 30, 40, 60 and 80 mM. In this plot an estimate of the Km is provided as the slope of the line representing the linear regression curve through the points and the Vmax by the intercept of the regression curve with the y axis.

Table 37 compares properties of a prokaryotic HpaI enzyme (*P. putida* HpaI) to that of HpaI-like enzyme of arabidopsis (At4g10750) soybean (Glyma09g21760) and rice (Os09g36030).

TABLE 37

Comparison of properties of a prokaryotic HpaI enzyme (*P. putida* HpaI) to that of HpaI-like enzyme of arabidopsis (At4g10750) soybeans (Glyma09g21760) and rice (Os09g36030).

enzyme/gene	pH optimum	preferred divalent ion
At4g10750	8	Mg2+
Glyma09g21760	8	Mg2+
Os09g36030	8.25	Mn2+
P putida HpaI	7.25	Co2+

enzyme/gene	v max (pmol s^{-1} mg^{-1} protein)	Kcat (s^{-1})	Km pyruvate (mM)	Km acetaldehyde (mM)
At4g10750	3139	0.6	3.97	1.37
Glyma09g21760	7853	1.6	13.18	1.74
Os09g36030	2104	0.4	7.49	1.48
P putida HpaI	122886	22.4	5.29	2.98

enzyme/gene	Kcat/Km (pyr)	Kcat/Km (acetaldehyde)
At4g10750	0.2	0.5
Glyma09g21760	0.1	0.9
Os09g36030	0.1	0.3
P putida HpaI	4.2	7.5

A DNA sequence encoding a variant of the processed, plastid localized At4g10750 protein in which arginine 83 (R83) is replaced by glycine (G83) was generated as follows: The n-terminal fragment of the gene was PCR amplified using PCR primers AthHpaIL fwd (SEQ ID NO:71) and AthHpaIL G83 rev (SEQ ID NO: 96) and plasmid DNA of pKR1478- At4g10750 (SEQ ID NO:53). The c-terminal fragment of the gene was PCR amplified using PCR primers AthHpaIL rev (SEQ ID NO:72) AthHpaIL G83 fwd (SEQ ID NO: 97) and plasmid DNA of pKR1478-At4g10750 (SEQ ID NO:53). PCR products of both reactions were combined and used as template in a PCR reactions with primers AthHpaIL fwd (SEQ ID NO:71) and AthHpaIL rev (SEQ ID NO:72). PCR products were cloned into pGEM®-T Easy to give pGEM®-T At4g10750-G83 (SEQ ID NO:98).

pGEM®-T At4g10750- G83 (SEQ ID NO:98) was digested with NcoI SalI. A restriction fragment of 890 bp was gel-purified and ligated to NcoI SalI-linearized plasmid DNA of pET28a to give pET28a At4g10750-G83 (SEQ ID NO: 99). The amino acid sequence of the At4g10750-G83 gene product including a c-terminal pET28A-derived hexa-histidine tag are set forth as SEQ ID NO:100 .

100 mL flasks each containing 25 of LB medium supplemented with 50 μg/mL kanamycin were inoculated with *E. coli* cells of strain Rosetta™(DE3)pLysS carrying either pET28a At4g10750 (SEQ ID NO:74) or pET28a At4g10750-G83 (SEQ ID NO:99). The cultures were grown at 37 °C until a cell density ($OD_{\lambda=600nm}$) of 0.6 was achieved. The cultures were cooled to 16 °C on ice. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was then added to a final concentration of 0.1 mM followed by continued culture at 16 °C for 36 h. From each culture duplicate sample of 1.5 mL were harvested by centrifugation and resuspended in 200 μL of 50 mM Hepes/KOH pH8. 20 μL of toluene were added to each cell suspension. Toluene-treated cell suspensions were incubated at 37 °C for 20 min. Aldol addition enzyme activity of toluene-treated cell suspensions was assayed as follows. Enzyme

assays consisted of 2 mM $MgCl_2$, 10 mM acetaldehyde, 10 mM pyruvate, 50 mM Hepes/KOH pH8 and 50 μL of toluene treated cell suspensions in a final volume of 100 μL. Enzyme assays were incubated at 27 °C for 20 min and stopped by addition of 20 μL of concentrated HCl followed by incubation at 100 °C. Enzyme assays were cleared by centrifugation and 4-hydroxy-2-oxovalerate was quantitated by HPLC as described above. Table 38 shows that that there is a 14-fold reduction of aldol addition activity in *E coli* cell suspensions transformed with pET28a At4g10750-G83 compared to *E. coli* cell suspensions transformed with pET28a At4g10750. SDS/PAGE analysis of protein extracts of both cultures showed similar levels of recombinantly produced protein.

TABLE 38

Aldol addition activity of E. coli cultures carrying pET28a At4g10750 or pET28a At4g10750-G83

sample	aldol addition activity (nmol mL^{-1} min^{-1})
pET28a At4g10750	26.5
pET28a At4g10750	26.7
pET28a At4g10750-G83	1.9
pET28a At4g10750-G83	1.8

EXAMPLE 26

Expression of plastid targeted, bacterial HpaI enzymes in developing seed

The following example describes DNA constructs for plastid-targeted expression of bacterial HpaI enzymes in developing seed. Transgenic plants generated with these DNA constructs have altered composition of seed storage compounds such as oil, protein and carbohydrate.

A DNA sequence encoding a signal sequence for plastid targeting was PCR-amplified from pKR1478- At4g10750 using primers HpaILORF FWD (SEQ ID NO:11) and FUSION REV (SEQ ID NO: 101)

to give PCR product 1. A DNA sequence encoding the *P. putida* HpaI protein and a pET29a-derived c-terminal 6xHIS tag was amplified from plasmid DNA of pET29a

HpaI PP (SEQ ID NO:94) using primers FUSION FWD: (SEQ ID NO: 102) and pET29a 3prime:

(SEQ ID NO: 103) to give PCR product 2. PCR products 1 and 2 were combined and used as template in a PCR reaction with HpaILORF FWD (SEQ ID NO:11) and pET29a 3prime (SEQ ID NO: 102). A PCR product of 1070 bp was extracted form agaroase gels and cloned into the pCR®8/GW/TOPO® vector (Invitrogen) to give pCR8GW- plastid HpaI PP (SEQ ID NO:104). The ORF comprised of DNA sequences encoding plastid targeting signal, *P. putida* HpaI enzyme and pET29a-derived 6x His tag 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 - plastid HpaI PP is set forth as SEQ ID NO:105. The fusion protein expressed by this plant transformation vector is set forth as SEQ ID NO:106. It is comprised of plastid targeting signal of At4g10750, the catalytic domain of HpaI of *Pseudomonas putida* (DSM 12585) and a c-terminal pET29a-derived hexa-histidine tag. The plasmid was used for agrobacterium-mediated transformation of Arabidopsis plants as described in Example 4. Seed oil content of wt control plants and T1 plants generated with plasmid pKR1478 - plastid HpaI PP can be measured by NMR as described in Example 2.

EXAMPLE 27

Composition of cDNA Libraries; Isolation and Sequencing of cDNA Clones

cDNA libraries representing mRNAs from various tissues of *Momordica charantia* (balsam pear), *Aclepsia syriaca* (milkweed), *and Tulipa gesneriana* (tulip) were prepared. The characteristics of the libraries are described below.

TABLE 39

cDNA Libraries from Tulip, milkweed, Balsam pear and mays

Library	Tissue	Clone
etp1c	Tulipa (Gesneriana, Apeldoorn)stage 3 pistil	etp1c.pk001.g3:fis etp1c.pk003.b22:fis
mas1c	developing fibers of common milkweed stage 1	mas1c.pk012.d9.f
fds1n	Balsam pear (Momordica charantia) developing	fds1n.pk007.i18

	seeds	
cfp2n	Maize Silk pollinated and unpollinated, pooled, Full-length enriched, normalized	cfp2n.pk070b11.fis1

cDNA libraries may be prepared by any one of many methods available. For example, the cDNAs may be introduced into plasmid vectors by first preparing the cDNA libraries in Uni-ZAP™ XR vectors according to the manufacturer's protocol (Stratagene Cloning Systems, La Jolla, CA). The Uni-ZAP™ XR libraries are converted into plasmid libraries according to the protocol provided by Stratagene. Upon conversion, cDNA inserts will be contained in the plasmid vector pBluescript. In addition, the cDNAs may be introduced directly into precut Bluescript II SK(+) vectors (Stratagene) using T4 DNA ligase (New England Biolabs), followed by transfection into DH10B cells according to the manufacturer's protocol (GIBCO BRL Products). Once the cDNA inserts are in plasmid vectors, plasmid DNAs are prepared from randomly picked bacterial colonies containing recombinant pBluescript plasmids, or the insert cDNA sequences are amplified via polymerase chain reaction using primers specific for vector sequences flanking the inserted cDNA sequences. Amplified insert DNAs or plasmid DNAs are sequenced in dye-primer sequencing reactions to generate partial cDNA sequences (expressed sequence tags or "ESTs"; see Adams et al., (1991) *Science 252*:1651-1656). The resulting ESTs are analyzed using a Perkin Elmer Model 377 fluorescent sequencer.

Full-insert sequence (FIS) data is generated utilizing a modified transposition protocol. Clones identified for FIS are recovered from archived glycerol stocks as single colonies, and plasmid DNAs are isolated via alkaline lysis. Isolated DNA templates are reacted with vector primed M13 forward and reverse oligonucleotides in a PCR-based sequencing reaction and loaded onto automated sequencers. Confirmation of clone identification is performed by sequence alignment to the original EST sequence from which the FIS request is made.

Confirmed templates are transposed via the Primer Island transposition kit (PE Applied Biosystems, Foster City, CA) which is based upon the *Saccharomyces cerevisiae* Ty1 transposable element (Devine and Boeke (1994) *Nucleic Acids Res. 22*:3765-3772). The *in vitro* transposition system places unique binding sites

randomly throughout a population of large DNA molecules. The transposed DNA is then used to transform DH10B electro-competent cells (Gibco BRL/Life Technologies, Rockville, MD) via electroporation. The transposable element contains an additional selectable marker (named DHFR; Fling and Richards (1983) *Nucleic Acids Res. 11*:5147-5158), allowing for dual selection on agar plates of only those subclones containing the integrated transposon. Multiple subclones are randomly selected from each transposition reaction, plasmid DNAs are prepared via alkaline lysis, and templates are sequenced (ABI Prism dye-terminator ReadyReaction mix) outward from the transposition event site, utilizing unique primers specific to the binding sites within the transposon.

Sequence data is collected (ABI Prism Collections) and assembled using Phred and Phrap (Ewing et al. (1998) *Genome Res. 8*:175-185; Ewing and Green (1998) *Genome Res. 8*:186-194). Phred is a public domain software program which re-reads the ABI sequence data, re-calls the bases, assigns quality values, and writes the base calls and quality values into editable output files. The Phrap sequence assembly program uses these quality values to increase the accuracy of the assembled sequence contigs. Assemblies are viewed by the Consed sequence editor (Gordon et al. (1998) *Genome Res. 8*:195-202).

In some of the clones the cDNA fragment corresponds to a portion of the 3'-terminus of the gene and does not cover the entire open reading frame. In order to obtain the upstream information one of two different protocols are used. The first of these methods results in the production of a fragment of DNA containing a portion of the desired gene sequence while the second method results in the production of a fragment containing the entire open reading frame. Both of these methods use two rounds of PCR amplification to obtain fragments from one or more libraries. The libraries some times are chosen based on previous knowledge that the specific gene should be found in a certain tissue and some times are randomly-chosen. Reactions to obtain the same gene may be performed on several libraries in parallel or on a pool of libraries. Library pools are normally prepared using from 3 to 5 different libraries and normalized to a uniform dilution. In the first round of amplification both methods use a vector-specific (forward) primer corresponding to a portion of the vector located at the 5'-terminus of the clone coupled with a gene-specific (reverse) primer. The first method uses a sequence that is

complementary to a portion of the already known gene sequence while the second method uses a gene-specific primer complementary to a portion of the 3'-untranslated region (also referred to as UTR). In the second round of amplification a nested set of primers is used for both methods. The resulting DNA fragment is ligated into a pBluescript vector using a commercial kit and following the manufacturer's protocol. This kit is selected from many available from several vendors including Invitrogen™ (Carlsbad, CA), Promega Biotech (Madison, WI), and Gibco-BRL (Gaithersburg, MD). The plasmid DNA is isolated by alkaline lysis method and submitted for sequencing and assembly using Phred/Phrap, as above.

EXAMPLE 28

Identification of cDNA Clones

cDNA clones encoding HpaI-like polypeptides were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol. 215*:403-410; see also the explanation of the BLAST algorithm on the world wide web site for the National Center for Biotechnology Information at the National Library of Medicine of the National Institutes of Health) searches for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences obtained as described in Example 6 were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish and States (1993) *Nat. Genet. 3*:266-272) provided by the NCBI. For convenience, the P-value (probability) of observing a match of a cDNA sequence to a sequence contained in the searched databases merely by chance as calculated by BLAST are reported herein as "pLog" values, which represent the negative of the logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the likelihood that the cDNA sequence and the BLAST "hit" represent homologous proteins.

ESTs submitted for analysis are compared to the Genbank database as described above. ESTs that contain sequences more 5- or 3-prime can be found by using the BLASTn algorithm (Altschul et al (1997) *Nucleic Acids Res.* *25*:3389-3402.) against the Du Pont proprietary database comparing nucleotide sequences that share common or overlapping regions of sequence homology. Where common or overlapping sequences exist between two or more nucleic acid fragments, the sequences can be assembled into a single contiguous nucleotide sequence, thus extending the original fragment in either the 5 or 3 prime direction. Once the most 5-prime EST is identified, its complete sequence can be determined by Full Insert Sequencing as described in Example 6. Homologous genes belonging to different species can be found by comparing the amino acid sequence of a known gene (from either a proprietary source or a public database) against an EST database using the tBLASTn algorithm. The tBLASTn algorithm searches an amino acid query against a nucleotide database that is translated in all 6 reading frames. This search allows for differences in nucleotide codon usage between different species, and for codon degeneracy.

EXAMPLE 29

Characterization of cDNA Clones Encoding HpaI-like Polypeptides

The BLASTX search using the EST sequences from clones listed in Table xx revealed similarity of the polypeptides encoded by the cDNAs to HpaI-like polypeptide from *Arabidopsis* (At4g10750) corresponding to SEQ ID NO's :47), Shown in Table 40 are the percent identities results for the proteins encoded by individual ESTs ("EST"), the sequences of the entire cDNA inserts comprising the indicated cDNA clones ("FIS"), the sequences of contigs assembled from two or more EST, FIS or PCR sequences ("Contig"), or sequences encoding an entire or functional protein derived from an FIS or a contig ("CGS"):

TABLE 40

Percent Identity for HpaI-like Polypeptides

Sequence	Status	NCBI GI No.	% identity
etp1c.pk001.g4:fis SEQ ID NO:120	FIS	255587508 (Ricinus communis) SEQ ID NO:109	53.6
etp1c.pk003.b22:fis SEQ ID NO:121	FIS	225426623 (Vitis vinifera) SEQ ID NO:111	53.4
mas1c.pk012.d9.f SEQ ID NO:122	FIS	225426623 (Vitis vinifera) SEQ ID NO:111	55.8
fds1n.pk007.i18 SEQ ID NO:123	CGS	225426623 (Vitis vinifera) SEQ ID NO:111	54.0
cfp2n.pk070b11.fis1 SEQ ID NO:147	CGS	226510158 (Zea mays) SEQ ID NO:33	99.7

Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was 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 using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

Sequence alignments and BLAST scores and probabilities indicate that the nucleic acid fragments comprising the instant cDNA clones encode HpaI-like polypeptides.

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 85% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 29, 31, 33, 35, 49, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, and 147 and wherein seed obtained from said transgenic plant has an altered i.e. increased or decreased oil, protein, starch and/or soluble carbohydrate content and/or altered seed weight 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 85% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 29, 31, 33, 35, 49, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, and 147 and wherein said transgenic seed has an altered oil, protein, starch and/or soluble carbohydrate content and/or altered seed weight 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 85% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 29, 31, 33, 35, 49, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, and 147and wherein said transgenic seed has an increased protein 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 85% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 29, 31, 33, 35, 49, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, and 123 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 85% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: : 29, 31, 33, 35, 49, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, and 123 , 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 85% 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 plastidic HpaIL aldolase polypeptide,

and wherein said plant has an altered oil, protein, and/or soluble carbohydrate content and/or altered seed weight 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: 28, 30, 32, 34, 48, 124, 125, 126,127, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, or 146;or

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

wherein (a) or (b) is of sufficient length to inhibit expression of endogenous plastidic HpaIL aldolase proteins or HpaIL aldolase 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, when compared to a seed from a control plant not comprising said recombinant DNA construct.

6. A transgenic seed comprising a recombinant DNA construct comprising:
   (a) all or part of the nucleotide sequence set forth in SEQ ID NO: 28, 30, 32, 34, 48, 124, 125, 126,127, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, or 146; or
   (b) the full-length complement of (a):

wherein (a) or (b) is of sufficient length to inhibit expression of endogenous plastidic HpaIL aldolase 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, when compared to a seed from a control plant not comprising said recombinant DNA construct.

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 85% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 29, 31, 33, 35, 49, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, and 123; and (b) regenerating a plant from the transformed plant cell and optionally (c) obtaining a progeny plant derived from the transgenic plant, wherein said progeny plant comprises in its genome the recombinant DNA construct and seed obtained from said progeny plant exhibit an altered i.e. increased or decreased oil, protein, starch and/or soluble carbohydrate content and/or altered seed weight when compared to a control plant not comprising said recombinant DNA construct.

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 85% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 29, 31, 33, 35, 49, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, and 123; 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 and/or altered seed weight, as when compared to a seed from a control plant not comprising said recombinant DNA construct.

9. 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 85% sequence identity, based on the Clustal V method of alignment, when compared to SEQ ID NO: 29, 31, 33, 35, 49, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, and 123; 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 prtoein content of at least 0.5 % when compared to a seed from a control plant not comprising said recombinant DNA construct.

10. 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: 28, 30, 32, 34, 48, 124, 125, 126,127, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, or 145; or

(ii) the full-length complement of (i);

wherein (i) or (ii) is of sufficient length to inhibit expression of endogenous plastidic HpaIL aldolase 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 and/or altered seed weight, when compared to a seed from a control plant not comprising said recombinant DNA construct.

11. 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: 28, 30, 32, 34, 48, 124, 125, 126,127, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, or 146; or

(ii) the full-length complement of (i);

wherein (i) or (ii) is of sufficient length to inhibit expression of endogenous plastidic HpaIL aldolase 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, when compared to a seed from a control plant not comprising said recombinant DNA construct.

12. A transgenic plant comprising at least one DGAT sequence and a construct downregulating plastidic Hpal or Hpal-like activity, wherein the DGAT sequence and the plastidic Hpal or Hpal-like construct can be in the same recombinant construct or in separate recombinant constructs, and wherein seed obtained from said transgenic plant has an increased oil content when compared to the oil content of seed obtained from a control plant not comprising said construct or when compared to transgenic seed obtained from a transgenic plant comprising either said DGAT sequences alone or said construct downregulating Hpal-like activity alone.

13.. A transgenic seed obtained from the transgenic plant of claim1 comprising at least one DGAT sequence and a construct downregulating Hpal or Hpal-like activity, wherein the DGAT sequence and the plastidic Hpal-like construct

can be in the same recombinant construct or in separate recombinant constructs and wherein the oil content of said transgenic seed is increased when compared to the oil content of control seed not comprising said construct or transgenic seed comprising either said DGAT sequence alone or said construct downregulating HpAI or HpaI-like activity alone.

14. The transgenic seed of any one of claim 1, 2, 3, 4, 5, 6, 12 or 13 wherein the transgenic seed is obtained from a monocot or dicot plant.

15. The transgenic seed of any one of claim 1, 2, 3, 4,5, 6, 12 or 13 wherein the transgenic seed is obtained from a maize or soybean plant.

16. The transgenic seed of any one of claim 1, 2, 3, 4, 5, 6, 12 or 13 wherein the at least one regulatory element is a seed-specific or seed-preferred promoter.

17. The transgenic seed of any one of claim 1, 2, 3, 4, 5, 6, 12 or 13 wherein at least one regulatory element is an endosperm or embryo-specific promoter.

18. A method for increasing the oil content of a seed comprising:

(a) transforming at least one cell with at least one recombinant construct having at least one DGAT sequence and a construct downregulating plastidic HpaI or HpaI-like activity wherein the DGAT sequence and the HpaI or HpaI-like construct can be in the same recombinant construct or in separate recombinant constructs;

(b) selecting the transformed soybean cell(s) of step (a) having an increased oil content when compared to the oil content of a control cell not comprising said construct or when compared to to transgenic seed obtained from a transgenic plant comprising either said DGAT sequences alone or said construct downregulating HpAI or HpaI-like activity alone.

19. The method of any one of claims 7, 8, 9, 10,11, or 18, wherein the transgenic seed is obtained from a transgenic dicot plant comprising in its genome the recombinant construct.

20. The method of any one of claims 7, 8, 9, 10,11, or 18, wherein the dicot plant is soybean.

20. Transgenic seed obtained by the method of any one of claims 7, 8, 9, 10, 11, or 18.

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

10.

22. The transgenic seed obtained by the method of claim 7, 8, 9, 10,11, or 18, wherein the transgenic seed is obtained from a monocot or dicot plant.

23. A product and/or by-product from transgenic seed of any one of claims 2 or 13 wherein the plant is maize or soybean.

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

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

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

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

28. An isolated polynucleotide comprising:

(a) a nucleotide sequence encoding a polypeptide with HpAIL aldolase activity, wherein, based on the Clustal V method of alignment with pairwise alignment default parameters of KTUPLE=1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5, the polypeptide has an amino acid sequence of at least 75% sequence identity when compared to SEQ ID NO:120, 121, 122 or 123, or

(b) the full complement of the nucleotide sequence of (a).

29. The polynucleotide of Claim 28, wherein the amino acid sequence of the polypeptide comprises SEQ ID NO: 120, 121, 122 or 123.

30. The polynucleotide of Claim 28 wherein the nucleotide sequence comprises SEQ ID NO:124, 125, 126, or 127.

31. A plant or seed comprising a recombinant DNA construct, wherein the recombinant DNA construct comprises the polynucleotide of any one of Claims 18 to 30 operably linked to at least one regulatory sequence.

32. An isolated polynucleotide encoding a polypeptide, wherein said polynucleotide is capable of altering the endogenous expression of plastidic Hpal-like activity and wherein said polypeptide comprises a chloroplast transit peptide and at least one motif selected from the group consisting of: SEQ ID NO: 128, 129, 130, 131, or 132.

33. An isolated polynucleotide encoding a plant Hpal-like polypeptide, wherein said polynucleotide is capable of altering the endogenous expression of plastidial Hpal-like activity and wherein said polypeptide has a Km (acetaldehyde) at least 1.7 fold lower than than the Km (acetaldehyde) of bacterial HpaIL aldolase activity and a Vmax of at least 15 fold lower than the bacterial HpaIL aldolase activity.

34. A recombinant DNA construct comprising the isolated polynucleotide of any aone of claims 32 and 33, further wherein a plant or seed comprising said recombinant DNA construct, has an altered i.e. increased or decreased oil, protein, starch and/or soluble carbohydrate content and/or altered seed weight when compared to a control plant not comprising said recombinant DNA construct.

35. A method of altering i.e. increasing or decreasing oil, protein, starch and/or coluble carbohydrate content and or altering seed weight, comprising:

a) transforming a plant with the recombinant DNA construct of claim 34;

b) growing the transformed plant under conditions suitable for the expression of the recombinant DNA construct; and

c) selecting those plant having altered i.e. increased or decreased oil, protein, starch and/or soluble carbohydrate content and/or altered seed weight when compared to a control plant not comprising said recombinant DNA construct.

36. A method to isolate nucleic acid fragments encoding plastidial Hpal-like polypeptides, comprising:

a) comparing SEQ ID NOs: 128, 129, 130, 131, or 132 with other polypeptide sequences encoding a plastidial Hpal-like polypeptides;

b) identifying the conserved sequences obtained in step (a);

c) making region-specific nucleotide probe(s) or oligomer(s) based on the conserved sequences identified in step (b); and

(d) using the nucleotide probe(s) or oligomer(s) of step (c) to isolate Hpal-like sequences;

e) selecting those sequences comprising a chloroplast transit peptide.

FIG. 1A

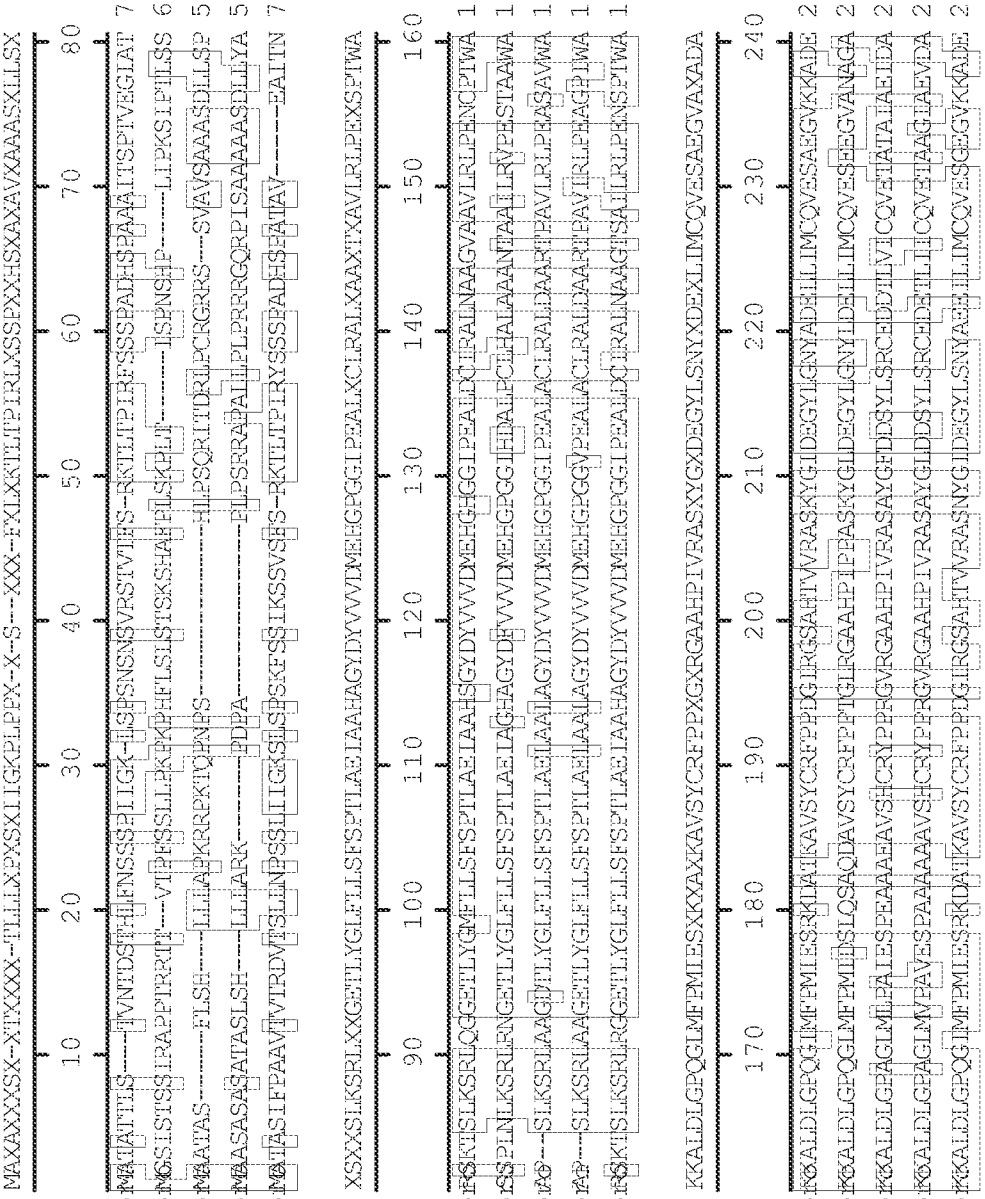


FIG.1B

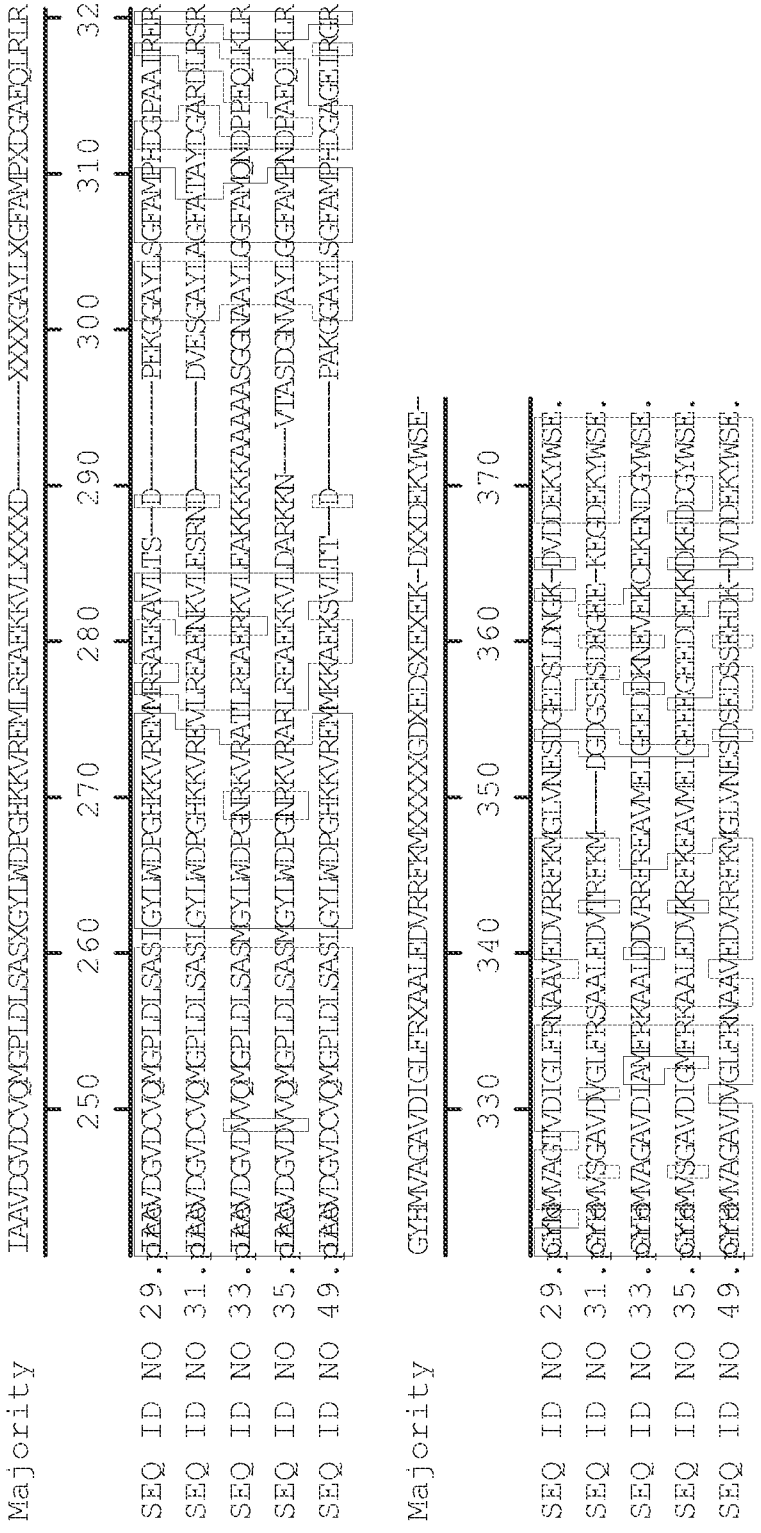

Majority
250
260
270
280
290
300
310
320
SEQ ID NO 29.
SEQ ID NO 31.
SEQ ID NO 33.
SEQ ID NO 35.
SEQ ID NO 49.
Majority
330
340
350
360
370
SEQ ID NO 29.
SEQ ID NO 31.
SEQ ID NO 33.
SEQ ID NO 35.
SEQ ID NO 49.

Fig.2

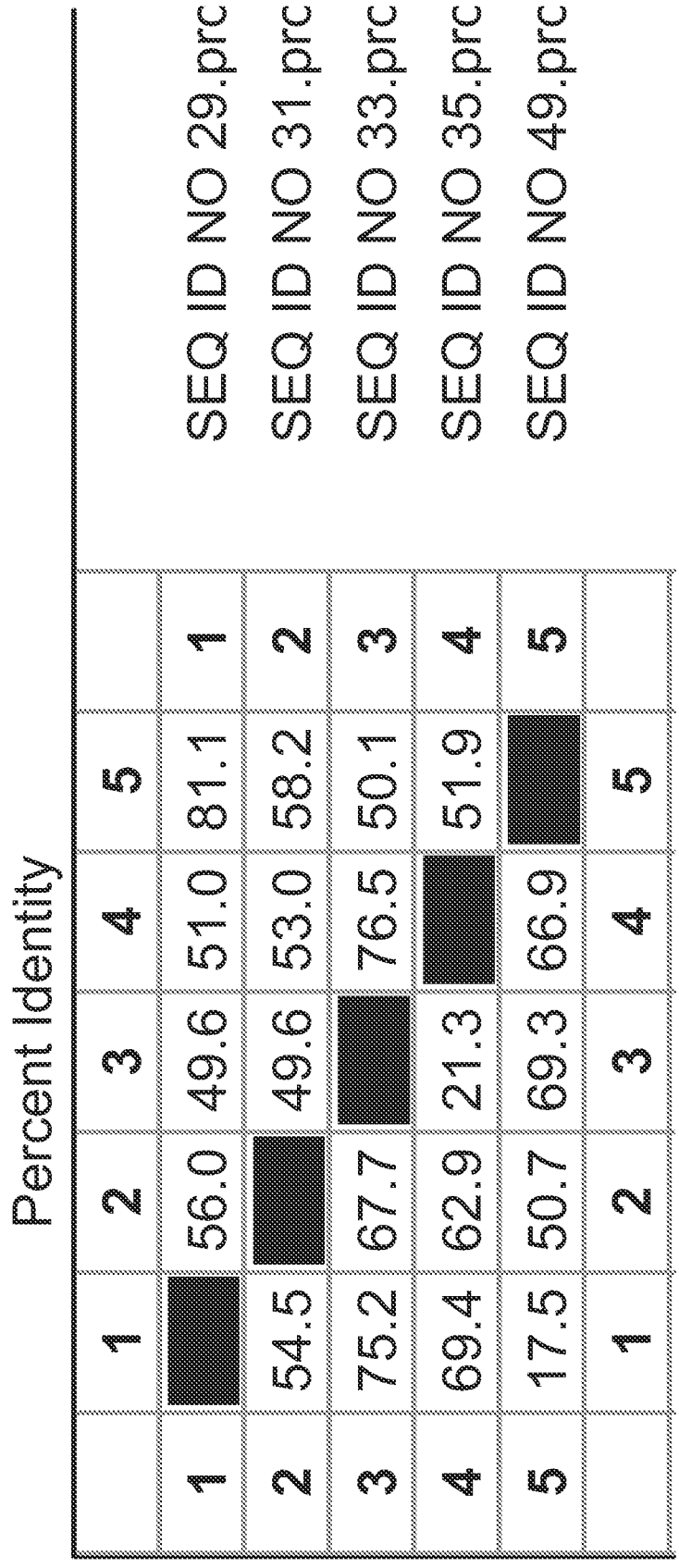

Percent Identity
1 2 3 4 5
1 56.0 49.6 51.0 81.1
2 54.5 49.6 53.0 58.2
3 75.2 67.7 76.5 50.1
4 69.4 62.9 21.3 51.9
5 17.5 50.7 69.3 66.9
1 2 3 4 5
SEQ ID NO 29.prc
SEQ ID NO 31.prc
SEQ ID NO 33.prc
SEQ ID NO 35.prc
SEQ ID NO 49.prc

FIG.3

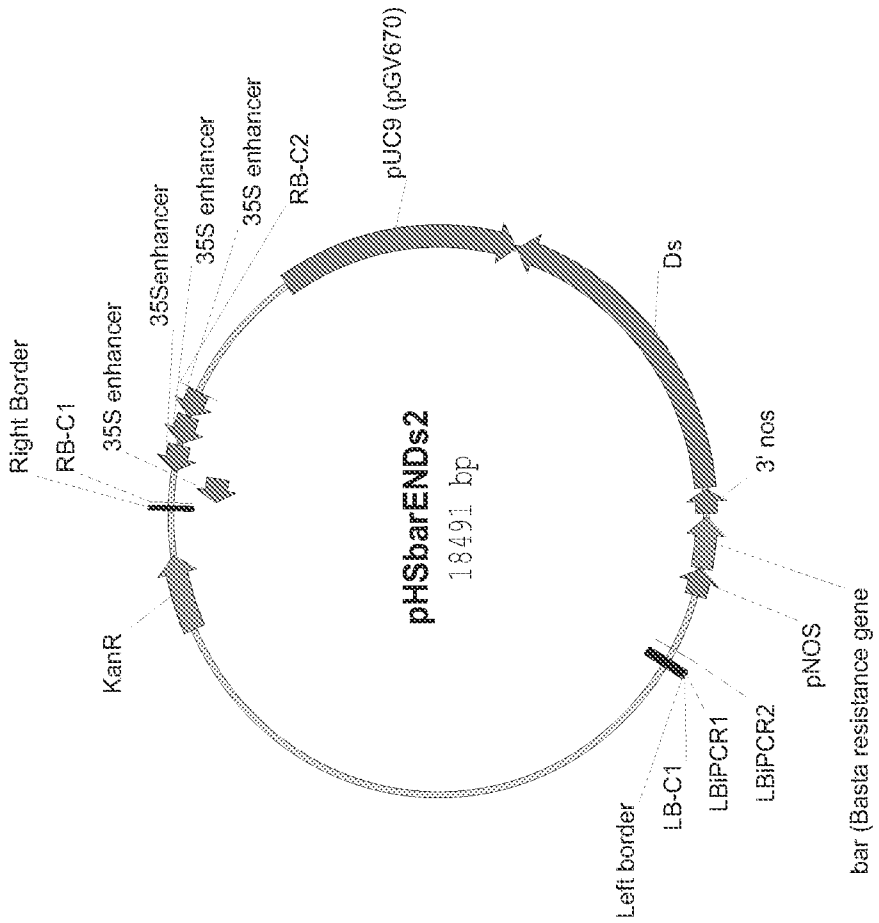

pHSbarENDs2
18491 bp
Right Border
RB-C1
35S enhancer
35Senhancer
35S enhancer
35S enhancer
RB-C2
pUC9 (pGV670)
Ds
3' nos
bar (Basta resistance gene
pNOS
LBiPCR2
LBiPCR1
LB-C1
Left border
KanR

FIG.4A


SEQ ID NO:107
SEQ ID NO-108
SEQ ID NO-109
SEQ ID NO-110
SEQ ID NO-111
SEQ ID NO-112
SEQ ID NO-113
SEQ ID NO-114
SEQ ID NO-115
SEQ ID NO-116
SEQ ID NO-117
SEQ ID NO-118
SEQ ID NO-119
SEQ ID NO-120
SEQ ID NO-121
SEQ ID NO 29
SEQ ID NO 31
SEQ ID NO 33
SEQ ID NO 35
SEQ ID NO 49
Motif I

FIG.4B

```
SEQ_ID_NO-118  YDYVVVDMEHGPGGIPEALSCLRALDAARTPAVLRLPEASAVWAKKALDLGPAGLMIPAV
SEQ_ID_NO-119  YDYVVVDMEHGPGGIPEALACLRALDAARTPAVLRLPEASPVWAKKALDLGPAGLMLPAV
SEQ_ID_NO-120  YDYVVIDMEHGPGGIPEAIACLRALAVTGTPAIIRVPELCAALAKKALDIGPQGIMFPMI
SEQ_ID_NO-121  YDFVVVDMEHGPGDIIDSLSSIRALAAAGTPAIIRVPEKSAAWAKKALDLGPQGIMFPMV
SEQ_ID_NO_29   YDYVVVDMEHGHGGIPEALDCIRALNAAGVAAVLRLPENCPTWAKKALDLGPQGIMFPMI
SEQ_ID_NO_31   YDFVVVDMEHGPGGIHDALPCLHALAAANTAAILRVPESTAAWAKKALDLGPQGLMFPMI
SEQ_ID_NO_33   YDYVVVDMEHGPGGIPEALACLRALDAARTPAVLRLPEASAVWAKKALDLGPAGLMLPAI
SEQ_ID_NO_35   YDYVVVDMEHGPGGVPEALACLRALDAARTPAVIRLPEAGPIWAKKALDLGPAGLMVPAV
SEQ_ID_NO_49   YDYVVVDMEHGPGGIPEALDCIRALNAAGTSAILRLPENSPTWAKKALDLGPQGIMFPMI
               ----------------------          ----------------------------
                    Motif II                        Δ         Motif III

               +    *  +** ** ** * ** *   **** **  + *+    ++ ++ * **    +
SEQ_ID_NO-107  ESRKDATKAVSYCRFPPDGIRGSAHTVVRASNYGIDEGYLSNYAEEILIMCQVESSEGVK
SEQ_ID_NO-108  DSPKDAKKAVSYCRFPPDGIRGSAHTVVRASNYGINEGYLSNYMEDLLIMCQVETVDAVK
SEQ_ID_NO-109  ESPKMAKKAVSYCRFPPEGIRGSAHTVVRASSYGIDEGYLSNYGDELLIMCQVESEEGVK
SEQ_ID_NO-110  DGPKSARKAVSYCRFPPNGVRGSAHTVVRASSYGIDEGYLSNYEDDLLIMCQVECVDGVK
SEQ_ID_NO-111  DSPKLARKAVSYCRFPPAGVRGSAHTVVRASSYGIDAGYLSNYEEELLIMCQVESEEAVK
SEQ_ID_NO-112  DSPKDAKKAVSYCHFPPKGIRGSAHTVVRASNYGVDEEYLSTYEEEQLIMCQVESEEGVK
SEQ_ID_NO-113  DSPEAAKEAVSYCRFPPSGVRGSAHTVVRASGYGIDEGYLSNYEEELLIMCQVESEEGVK
SEQ_ID_NO-114  DSPEAAKEAVSYCRFPPSGVRGSAHTVVRASGYGIDEGYLSNYEEELLIMCQVESEEGVK
SEQ_ID_NO-115  ESPAAAAAAVSHCRYPPRGVRGAAHPIVRASVYGLDDSYLSRCEDDTLIICQVETAAGIA
SEQ_ID_NO-116  ESPAAAAEAVSHCRYPPRGVRGAAYPIVRASAYGLDDSYVSRCEDDTLIICQVETAAGVA
SEQ_ID_NO-117  ESPAAAAEAVSHCRYPPRGVRGAAHPIVRASAYGFDDSYISRCEDDTLVICQVETATGIA
SEQ_ID_NO-118  ESPAAAAEAVSYXRYPPRGVRGAAHPIVRASAYGLDDSYLSRCEDDTLLICQVETAAGIA
SEQ_ID_NO-119  ESPAAAAEAVSYCRYPPRGVRGAAHTVVRASAYGLDDSYLSRCEDETLIMCQVETAAGIA
SEQ_ID_NO-120  ESADQAELAVSYCRYPPKGIRGAAN-VVRASAYGFDEGYLKWCREELIVFCQVESEAGVA
SEQ_ID_NO-121  ENAKTAEQLVSYCRYPPRGIRGTAHVMVRASGYGTDDTYVHRCEEDLLILCMVETETGVD
SEQ_ID_NO_29   ESRKDATKAVSYCRFPPDGIRGSAHTVVRASKYGIDEGYLGNYADELLIMCQVESAEGVK
SEQ_ID_NO_31   DSLQSAQDAVSYCRFPPTGLRGAAHPIPPASKYGLDEGYLGNYLDELLIMCQVESEEGVA
SEQ_ID_NO_33   ESPEAAAEAVSHCRYPPRGVRGAAHPIVRASAYGFDDSYLSRCEDDTLVICQVETATAIA
SEQ_ID_NO_35   ESPAAAAAAVSHCRYPPRGVRGAAHPIVRASAYGLDDSYLSRCEDETLIICQVETAAGIA
SEQ_ID_NO_49   ESRKDATKAVSYCRFPPDGIRGSAHTVVRASNYGIDEGYLSNYAEEILIMCQVESGEGVK
               ----------------------------------------------------------
                                                Motif IV

                   ***++**  +*+** *+ **+* * +*           **
SEQ_ID_NO-107  KADEIAAVDGVDCVQMGPLDLSASLGYLWDPGHKKVREMMKKAEKSVLTSDP---------
SEQ_ID_NO-108  KVEEIAAVDGVDCVQMGPLDLSASLGYLWDPGHKNVREMLRTAERGVLKSDP---------
SEQ_ID_NO-109  KADEIAAVDGVDCIQMGPLDLSASMGYLWDPGHKKVREMMRAAEKAVLGSK-K--------
SEQ_ID_NO-110  KIDEIAAVEGVDCIQMGPLDLSASLGYLWDPGNKKVKEMMNTAEKGALKKKPL--------
SEQ_ID_NO-111  KIEDIAAVDGVDCIQMGPMDLSASMGYLWDPGNKKVREMMKVAEKGVLGTKKK--------
SEQ_ID_NO-112  KVEDIAAVEGVDCIQMGPLDLSASMGYLWDPGHKKVREMMGVAEKRVLGTKP---------
SEQ_ID_NO-113  RAEDIAAVAGVDCVQMGPLDLSASMGYLWDPGHRKVREMMRVAEKGVLG------------
SEQ_ID_NO-114  RAEDIAAVDGVDCVQMGPLDLSASMGYLWDPGHRKVREMMRVAEKGVLG------------
SEQ_ID_NO-115  EVDAIAAVDGVDVVQMGPLDLSASMGYLWDPGNRKVRARLREAEKKVLDARKKK---VTA
SEQ_ID_NO-116  EVDAIAAVDGVDVVQMGPLDLSASMGYLWDPGNRKVRAALREAERKVLEARKKK---VTA
SEQ_ID_NO-117  EIDAIAAIDGVDVVQMGPLDLSASMGYLWDPGNRKVRAALREAERKVLEARKKK--VAAA
SEQ_ID_NO-118  EVDAIASVDGVDVVQMGPLDMSASMGYLWDPGNRKVRAALREAERKVLEARKK---VAAP
SEQ_ID_NO-119  EVEAIAAVDGVDVVQMGPLDLSASMGYLWDPGNRKVRATLREAERKVLAARKKKDA---S
SEQ_ID_NO-120  ELDAIAAVDGVGGIQIGPRDLSASMGCLEEPDNPAVMETLRTAERRALAAGKKSTG----
SEQ_ID_NO-121  NIEEIASVDGVDGIQMGPRDLGASMGYLKDPGNPKARKVLLAAEKRVVALQK---------
SEQ_ID_NO_29   KADEIAAVDGVDCVQMGPLDLSASIGYLWDPGHKKVREMMRRAEKAVLTSDP---------
SEQ_ID_NO_31   NAGAIAAVDGVDCVQMGPLDLSASLGYLWDPGHKKVREVLREAENKVLESRND--------
SEQ_ID_NO_33   EIDAIAAVDGVDVVQMGPLDLSASMGYLWDPGNRKVRATLREAERKVLEAKKKKKAAAAA
SEQ_ID_NO_35   EVDAIAAVDGVDVVQMGPLDLSASMGYLWDPGNRKVRARLREAEKKVLDARKKN---VTA
SEQ_ID_NO_49   KADEIAAVDGVDCVQMGPLDLSASLGYLWDPGHKKVREMMKKAEKSVLTTDP---------
               ----------------------------
                       Motif V
```

FIG.4C

```
                          * *         *** ++ *  +   + + *** ++  *
SEQ_ID_NO-107 A-KGGAYLSGFAMPHDGAVEIRGRGYHMVAGAVDVGLFRNAAVEDVRRFKMGLVNESDGE
SEQ_ID_NO-108 G-DGGAFLAGFAMPHDPPVELGRRGYHMVSGAVDFALFRNAALADVKSFKNSVTVGFDDD
SEQ_ID_NO-109 G-GGGAYLAGFSMPHDGPIDLKSRGYNMVAGTVDVGLFRSAAVDDVKKFKMSLVQGSDDE
SEQ_ID_NO-110 D--GGAYLSGFAMPHDSPENLKSRGYHMVSGAVDIALFRNAAVEDVNKFKMSLDKGFEDQ
SEQ_ID_NO-111 P-REGAYLAGFAMPHDSPDDLRSRGYHMVSGAVDVGLFRSAAVEDVKKFKMGLKEASDDE
SEQ_ID_NO-112 G-NGGAYLAGFAMPHDGPDDLRARGYHMVSGAVDVGMFRNAAVEDVRRFKMGLMQGSDDE
SEQ_ID_NO-113 --GGKAYLAGFAMPHDAPLEMKSRGYHMVSGAVDVGLFRNAAVEDVARFKMNLTDDADDM
SEQ_ID_NO-114 --GGKAYLAGFAMPHDEPLEMKSRGYHMVVGAVELGLVKNAAVKEVARFIKNLNEKCQII
SEQ_ID_NO-115 SDGNVAYLGGFAMPNDPAEQLKLRGYHMVAGAADIGMFRKAALEDVKRFKEAVME--IGE
SEQ_ID_NO-116 SAGNAAYLGGFAMPNDPPEQLKMRGYHMVAGAVDIGLFRKAALEDVKWFKEAVME--IGE
SEQ_ID_NO-117 SDGNAAYLGGFAMQNDPPEQLKLRGYHMVAGAVDIAMFRKAALDDVKRFREAVME--IGE
SEQ_ID_NO-118 SDGNAAYLGGFAMQNDPPEQLKLRGYHMVAGAVDIGMFRKAALDDVKRFREAVME--IGE
SEQ_ID_NO-119 SDGNAAYLGGFAMPNDQAEQLKLRGYHMVAGAVDVGLFRKAALDDIKRFREAVME--IGE
SEQ_ID_NO-120 -----PYLAGMATALDSPGELLRRGYHMVGSGADLSLFRKAALQDVESFQKAKLAAATGD
SEQ_ID_NO-121 -AGKGPFLGGISTQQHKPEELRERGYNIIAGGVDVGLFRQAAVEDVQRFRRGAATPANKN
SEQ_ID_NO_29  E-KGGAYLSGFAMPHDGPAAIRERGYNMVAGTVDIGLFRNAAVEDVRRFKMGLVNESDGE
SEQ_ID_NO_31  DVESGAYLAGFATAYDGARDLRSRGYHMVSGAVDVGLFRSAALEDVTRFKM----DGDGS
SEQ_ID_NO_33  SGGNAAYLGGFAMQNDPPEQLKLRGYHMVAGAVDIAMFRKAALDDVRRFREAVME--IGE
SEQ_ID_NO_35  SDGNVAYLGGFAMPNDPAEQLKLRGYHMVSGAVDIGMFRKAALEDVKRFKEAVME--IGE
SEQ_ID_NO_49  A-KGGAYLSGFAMPHDGAGEIRGRGYHMVAGAVDVGLFRNAAVEDVRRFKMGLVNESDSE
                                          ----------------------------
                                                  Motif VI

SEQ_ID_NO-107 DSLEHDKDV---DDEKYWS-E------
SEQ_ID_NO-108 G--EEDKDG----DEKYWSE-------
SEQ_ID_NO-109 E--EHDKDA----DEKYWSE-------
SEQ_ID_NO-110 K--DH-KDG---E-EKYWSE-------
SEQ_ID_NO-111 K--ENAK------EDEKYWSE-------
SEQ_ID_NO-112 K--ERGRDG---EDEKYWSE-------
SEQ_ID_NO-113 T---NIMMM---LGGKGLRC-------
SEQ_ID_NO-114 K---KIKLG----GERIKKS-------
SEQ_ID_NO-115 EGDEYGDEKKDKEDDGYWS-E------
SEQ_ID_NO-116 EEGEEEDEK----DDGYWSE-------
SEQ_ID_NO-117 EGDEDEKD--EKENDGYWS-E------
SEQ_ID_NO-118 EGDDEEDEKDEKQNDGYWSE-------
SEQ_ID_NO-119 EGDEEEDEKLEKEADGYWSE-------
SEQ_ID_NO-120 EGVKERHTSNGVVKNRSYGSQMSNGY
SEQ_ID_NO-121 HNIVNGVHV----PSA------A------
SEQ_ID_NO_29  DSLDNGKDV---DDEKYWSE.------
SEQ_ID_NO_31  ESDEGEEKE---GDEKYWSE.------
SEQ_ID_NO_33  EDDKNEVEKCEKENDGYWSE.------
SEQ_ID_NO_35  EEGEEDDEKKDKEDDGYWSE.------
SEQ_ID_NO_49  DSSEHDKDV---DDEKYWSE.------
```