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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 November 2004 (04.11.2004)

PCT

(10) International Publication Number
WO 2004/093528 A2

(51) International Patent Classification[7]: A01H

(21) International Application Number:
PCT/US2004/012070

(22) International Filing Date: 19 April 2004 (19.04.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/464,550 22 April 2003 (22.04.2003) US

(71) Applicant *(for all designated States except US)*: **AGRINOMICS LLC** [US/US]; 16160 S.W. Upper Boones Ferry Road, Portland, OR 97224-7744 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **LIGHTNER, Jonathan** [US/US]; 28750 S. Heisinger Lane, Mulino, OR 97042 (US). **NG, Hein, Tsoeng** [CN/US]; 12975 SW 19th Street, Beaverton, OR 97008 (US).

(74) Agents: **SALE, Terri, L.** et al.; P.O. Box 511, 170 Harbor Way, South San Francisco, CA 94083-0511 (US).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, 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*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: GENERATION OF PLANTS WITH ALTERED OIL CONTENT

(57) Abstract: The present invention is directed to plants that display an altered oil content phenotype due to altered expression of a HIO105.1 nucleic acid. The invention is further directed to methods of generating plants with an altered oil content phenotype.

WO 2004/093528 A2

GENERATION OF PLANTS WITH ALTERED OIL CONTENT

REFERENCE TO RELATED APPLICATIONS

This application claims priority to U.S. provisional patent application 60/464,550 filed 4/22/2003, the contents of which are hereby incorporated by reference

BACKGROUND OF THE INVENTION

The ability to manipulate the composition of crop seeds, particularly the content and composition of seed oils, has important applications in the agricultural industries, relating both to processed food oils and to oils for animal feeding. Seeds of agricultural crops contain a variety of valuable constituents, including oil, protein and starch. Industrial processing can separate some or all of these constituents for individual sale in specific applications. For instance, nearly 60% of the US soybean crop is crushed by the soy processing industry. Soy processing yields purified oil, which is sold at high value, while the remainder is sold principally for lower value livestock feed (US Soybean Board, 2001 Soy Stats). Canola seed is crushed to produce oil and the co-product canola meal (Canola Council of Canada). Nearly 20% of the 1999/2000 US corn crop was industrially refined, primarily for production of starch, ethanol and oil (Corn Refiners Association). Thus, it is often desirable to maximize oil content of seeds. For instance, for processed oilseeds such as soy and canola, increasing the absolute oil content of the seed will increase the value of such grains. For processed corn it may be desired to either increase or decrease oil content, depending on utilization of other major constituents. Decreasing oil may improve the quality of isolated starch by reducing undesired flavors associated with oil oxidation. Alternatively, in ethanol production, where flavor is unimportant, increasing oil content may increase overall value. In many fed grains, such as corn and wheat, it is desirable to increase seed oil content, because oil has higher energy content than other seed constituents such as carbohydrate. Oilseed processing, like most grain processing businesses, is a capital-intensive business; thus small shifts in the distribution of products from the low valued components to the high value oil component can have substantial economic impacts for grain processors.

Biotechnological manipulation of oils can provide compositional alteration and improvement of oil yield. Compositional alterations include high oleic soybean and corn oil (US Pat Nos 6,229,033 and 6,248,939), and laurate-containing seeds (US Pat No 5,639,790), among others. Work in compositional alteration has predominantly focused

on processed oilseeds but has been readily extendable to non-oilseed crops, including corn. While there is considerable interest in increasing oil content, the only currently practiced biotechnology in this area is High-Oil Corn (HOC) technology (DuPont, US PAT NO: 5,704,160). HOC employs high oil pollinators developed by classical selection breeding along with elite (male-sterile) hybrid females in a production system referred to as TopCross. The TopCross High Oil system raises harvested grain oil content in maize from ~3.5% to ~7%, improving the energy content of the grain.

While it has been fruitful, the HOC production system has inherent limitations. First, the system of having a low percentage of pollinators responsible for an entire field's seed set contains inherent risks, particularly in drought years. Second, oil contents in current HOC fields have plateaued at about 9% oil. Finally, high-oil corn is not primarily a biochemical change, but rather an anatomical mutant (increased embryo size) that has the indirect result of increasing oil content. For these reasons, an alternative high oil strategy, particularly one that derives from an altered biochemical output, would be especially valuable.

The most obvious target crops for the processed oil market are soy and rapeseed, and a large body of commercial work (e.g., US Pat No: 5,952,544; PCT application WO9411516) demonstrates that *Arabidopsis* is an excellent model for oil metabolism in these crops. Biochemical screens of seed oil composition have identified *Arabidopsis* genes for many critical biosynthetic enzymes and have led to identification of agronomically important gene orthologs. For instance, screens using chemically mutagenized populations have identified lipid mutants whose seeds display altered fatty acid composition (Lemieux *et al.*, 1990; James and Dooner, 1990). T-DNA mutagenesis screens (Feldmann *et al.*, 1989) that detected altered fatty acid composition identified the omega 3 desaturase (*FAD3*) and delta-12 desaturase (*FAD2*) genes (US Pat No 5952544; Yadav *et al.*, 1993; Okuley *et al.*, 1994). A screen which focused on oil content rather than oil quality, analyzed chemically-induced mutants for wrinkled seeds or altered seed density, from which altered seed oil content was inferred (Focks and Benning, 1998). Another screen, designed to identify enzymes involved in production of very long chain fatty acids, identified a mutation in the gene encoding a diacylglycerol acyltransferase (DGAT) as being responsible for reduced triacyl glycerol accumulation in seeds (Katavic V *et al.*, 1995). It was further shown that seed-specific over-expression of the DGAT cDNA was associated with increased seed oil content (Jako *et al.*, 2001).

Activation tagging in plants refers to a method of generating random mutations by insertion of a heterologous nucleic acid construct comprising regulatory sequences (e.g., an enhancer) into a plant genome. The regulatory sequences can act to enhance transcription of one or more native plant genes; accordingly, activation tagging is a fruitful method for generating gain-of-function, generally dominant mutants (see, *e.g.*, Hayashi *et al.*, 1992; Weigel D *et al.* 2000). The inserted construct provides a molecular tag for rapid identification of the native plant whose mis-expression causes the mutant phenotype. Activation tagging may also cause loss-of-function phenotypes. The insertion may result in disruption of a native plant gene, in which case the phenotype is generally recessive.

Activation tagging has been used in various species, including tobacco and *Arabidopsis*, to identify many different kinds of mutant phenotypes and the genes associated with these phenotypes (Wilson *et al.*, 1996, Schaffer *et al.*, 1998, Fridborg *et al.*, 1999; Kardailsky *et al.*, 1999; Christensen S *et al.*, 1998).

SUMMARY OF THE INVENTION

The invention provides a transgenic plant having a high oil phenotype. The transgenic plant comprises a transformation vector comprising a nucleotide sequence that encodes or is complementary to a sequence that encodes a HIO105.1 polypeptide. In preferred embodiments, the transgenic plant is selected from the group consisting of rapeseed, soy, corn, sunflower, cotton, cocoa, safflower, oil palm, coconut palm, flax, castor and peanut. The invention further provides a method of producing oil comprising growing the transgenic plant and recovering oil from said plant.

The transgenic plant of the invention is produced by a method that comprises introducing into progenitor cells of the plant a plant transformation vector comprising a nucleotide sequence that encodes or is complementary to a sequence that encodes a HIO105.1 polypeptide, and growing the transformed progenitor cells to produce a transgenic plant, wherein the HIO105.1 polynucleotide sequence is expressed causing the high oil phenotype.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

Unless otherwise indicated, all technical and scientific terms used herein have the same meaning as they would to one skilled in the art of the present invention. Practitioners are particularly directed to Sambrook *et al.*, 1989, and Ausubel FM *et al.*,

1993, for definitions and terms of the art. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary.

As used herein, the term "vector" refers to a nucleic acid construct designed for transfer between different host cells. An "expression vector" refers to a vector that has the ability to incorporate and express heterologous DNA fragments in a foreign cell. Many prokaryotic and eukaryotic expression vectors are commercially available. Selection of appropriate expression vectors is within the knowledge of those having skill in the art.

A "heterologous" nucleic acid construct or sequence has a portion of the sequence that is not native to the plant cell in which it is expressed. Heterologous, with respect to a control sequence refers to a control sequence (*i.e.* promoter or enhancer) that does not function in nature to regulate the same gene the expression of which it is currently regulating. Generally, heterologous nucleic acid sequences are not endogenous to the cell or part of the genome in which they are present, and have been added to the cell, by infection, transfection, microinjection, electroporation, or the like. A "heterologous" nucleic acid construct may contain a control sequence/DNA coding sequence combination that is the same as, or different from a control sequence/DNA coding sequence combination found in the native plant.

As used herein, the term "gene" means the segment of DNA involved in producing a polypeptide chain, which may or may not include regions preceding and following the coding region, *e.g.* 5' untranslated (5' UTR) or "leader" sequences and 3' UTR or "trailer" sequences, as well as intervening sequences (introns) between individual coding segments (exons) and non-transcribed regulatory sequence.

As used herein, "recombinant" includes reference to a cell or vector, that has been modified by the introduction of a heterologous nucleic acid sequence or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found in identical form within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all as a result of deliberate human intervention.

As used herein, the term "gene expression" refers to the process by which a polypeptide is produced based on the nucleic acid sequence of a gene. The process includes both transcription and translation; accordingly, "expression" may refer to either a polynucleotide or polypeptide sequence, or both. Sometimes, expression of a polynucleotide sequence will not lead to protein translation. "Over-expression" refers to

increased expression of a polynucleotide and/or polypeptide sequence relative to its expression in a wild-type (or other reference [e.g., non-transgenic]) plant and may relate to a naturally-occurring or non-naturally occurring sequence. "Ectopic expression" refers to expression at a time, place, and/or increased level that does not naturally occur in the non-altered or wild-type plant. "Under-expression" refers to decreased expression of a polynucleotide and/or polypeptide sequence, generally of an endogenous gene, relative to its expression in a wild-type plant. The terms "mis-expression" and "altered expression" encompass over-expression, under-expression, and ectopic expression.

The term "introduced" in the context of inserting a nucleic acid sequence into a cell, means "transfection", or "transformation" or "transduction" and includes reference to the incorporation of a nucleic acid sequence into a eukaryotic or prokaryotic cell where the nucleic acid sequence may be incorporated into the genome of the cell (for example, chromosome, plasmid, plastid, or mitochondrial DNA), converted into an autonomous replicon, or transiently expressed (for example, transfected mRNA).

As used herein, a "plant cell" refers to any cell derived from a plant, including cells from undifferentiated tissue (*e.g.*, callus) as well as plant seeds, pollen, progagules and embryos.

As used herein, the terms "native" and "wild-type" relative to a given plant trait or phenotype refers to the form in which that trait or phenotype is found in the same variety of plant in nature.

As used herein, the term "modified" regarding a plant trait, refers to a change in the phenotype of a transgenic plant relative to the similar non-transgenic plant. An "interesting phenotype (trait)" with reference to a transgenic plant refers to an observable or measurable phenotype demonstrated by a T1 and/or subsequent generation plant, which is not displayed by the corresponding non-transgenic (i.e., a genotypically similar plant that has been raised or assayed under similar conditions). An interesting phenotype may represent an improvement in the plant or may provide a means to produce improvements in other plants. An "improvement" is a feature that may enhance the utility of a plant species or variety by providing the plant with a unique and/or novel quality. An "altered oil content phenotype" refers to measurable phenotype of a genetically modified plant, where the plant displays a statistically significant increase or decrease in overall oil content (i.e., the percentage of seed mass that is oil), as compared to the similar, but non-modified plant. A high oil phenotype refers to an increase in overall oil content.

As used herein, a "mutant" polynucleotide sequence or gene differs from the corresponding wild type polynucleotide sequence or gene either in terms of sequence or expression, where the difference contributes to a modified plant phenotype or trait. Relative to a plant or plant line, the term "mutant" refers to a plant or plant line which has a modified plant phenotype or trait, where the modified phenotype or trait is associated with the modified expression of a wild type polynucleotide sequence or gene.

As used herein, the term "T1" refers to the generation of plants from the seed of T0 plants. The T1 generation is the first set of transformed plants that can be selected by application of a selection agent, *e.g.*, an antibiotic or herbicide, for which the transgenic plant contains the corresponding resistance gene. The term "T2" refers to the generation of plants by self-fertilization of the flowers of T1 plants, previously selected as being transgenic. T3 plants are generated from T2 plants, etc. As used herein, the "direct progeny" of a given plant derives from the seed (or, sometimes, other tissue) of that plant and is in the immediately subsequent generation; for instance, for a given lineage, a T2 plant is the direct progeny of a T1 plant. The "indirect progeny" of a given plant derives from the seed (or other tissue) of the direct progeny of that plant, or from the seed (or other tissue) of subsequent generations in that lineage; for instance, a T3 plant is the indirect progeny of a T1 plant.

As used herein, the term "plant part" includes any plant organ or tissue, including, without limitation, seeds, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen, and microspores. Plant cells can be obtained from any plant organ or tissue and cultures prepared therefrom. The class of plants which can be used in the methods of the present invention is generally as broad as the class of higher plants amenable to transformation techniques, including both monocotyledenous and dicotyledenous plants.

As used herein, "transgenic plant" includes a plant that comprises within its genome a heterologous polynucleotide. The heterologous polynucleotide can be either stably integrated into the genome, or can be extra-chromosomal. Preferably, the polynucleotide of the present invention is stably integrated into the genome such that the polynucleotide is passed on to successive generations. A plant cell, tissue, organ, or plant into which the heterologous polynucleotides have been introduced is considered "transformed", "transfected", or "transgenic". Direct and indirect progeny of transformed plants or plant cells that also contain the heterologous polynucleotide are also considered transgenic.

Identification of Plants with an Altered Oil Content Phenotype

We used an *Arabidopsis* activation tagging screen to identify the association between the gene we have designated "HIO105.1," (At2g40040; GI#18405244) encoding a non-secretory protein that lacks a transmembrane domain (GI#15225617), and an altered oil content phenotype (specifically, a high oil phenotype). Briefly, and as further described in the Examples, a large number of *Arabidopsis* plants were mutated with the pSKI015 vector, which comprises a T-DNA from the Ti plasmid of *Agrobacterium tumifaciens*, a viral enhancer element, and a selectable marker gene (Weigel *et al.*, 2000). When the T-DNA inserts into the genome of transformed plants, the enhancer element can cause up-regulation genes in the vicinity, generally within about 10 kilobase (kb) of the insertion. T1 plants were exposed to the selective agent in order to specifically recover transformed plants that expressed the selectable marker and therefore harbored T-DNA insertions. Samples of approximately 15-20 T2 seeds were collected from transformed T1 plants, and lipids were extracted from whole seeds. Gas chromatography (GC) analysis was performed to determine fatty acid content and composition of seed samples.

An *Arabidopsis* line that showed a high-oil phenotype was identified wherein oil content (i.e., fatty acids) constituted approximately 41.0% relative to a planting day average oil content of 37.1% (110% of PDA). The association of the HIO105.1 gene with the high oil phenotype was discovered by analysis of the genomic DNA sequence flanking the T-DNA insertion in the identified line. Accordingly, HIO105.1 genes and/or polypeptides may be employed in the development of genetically modified plants having a modified oil content phenotype ("a HIO105.1 phenotype"). HIO105.1 genes may be used in the generation of oilseed crops that provide improved oil yield from oilseed processing and in the generation of feed grain crops that provide increased energy for animal feeding. HIO105.1 genes may further be used to increase the oil content of specialty oil crops, in order to augment yield of desired unusual fatty acids. Transgenic plants that have been genetically modified to express HIO105.1 can be used in the production of oil, wherein the transgenic plants are grown, and oil is obtained from plant parts (e.g. seed) using standard methods.

HIO105.1 Nucleic Acids and Polypeptides

Arabidopsis HIO105.1 nucleic acid (genomic DNA) sequence is provided in SEQ ID NO:1 and in Genbank entry GI#18405244. The corresponding protein sequence is

provided in SEQ ID NO:2 and in GI#15225617. Nucleic acids and/or proteins that are orthologs or paralogs of *Arabidopsis* HIO105.1, are described in Example 3 below.

As used herein, the term "HIO105.1 polypeptide" refers to a full-length HIO105.1 protein or a fragment, derivative (variant), or ortholog thereof that is "functionally active," meaning that the protein fragment, derivative, or ortholog exhibits one or more or the functional activities associated with the polypeptide of SEQ ID NO:2. In one preferred embodiment, a functionally active HIO105.1 polypeptide causes an altered oil content phenotype when mis-expressed in a plant. In a further preferred embodiment, mis-expression of the HIO105.1 polypeptide causes a high oil phenotype in a plant. In another embodiment, a functionally active HIO105.1 polypeptide is capable of rescuing defective (including deficient) endogenous HIO105.1 activity when expressed in a plant or in plant cells; the rescuing polypeptide may be from the same or from a different species as that with defective activity. In another embodiment, a functionally active fragment of a full length HIO105.1 polypeptide (i.e., a native polypeptide having the sequence of SEQ ID NO:2 or a naturally occurring ortholog thereof) retains one of more of the biological properties associated with the full-length HIO105.1 polypeptide, such as signaling activity, binding activity, catalytic activity, or cellular or extra-cellular localizing activity. A HIO105.1 fragment preferably comprises a HIO105.1 domain, such as a C- or N-terminal or catalytic domain, among others, and preferably comprises at least 10, preferably at least 20, more preferably at least 25, and most preferably at least 50 contiguous amino acids of a HIO105.1 protein. Functional domains can be identified using the PFAM program (Bateman A *et al.*, 1999 Nucleic Acids Res 27:260-262; website at pfam.wustl.edu). Functionally active variants of full-length HIO105.1 polypeptides or fragments thereof include polypeptides with amino acid insertions, deletions, or substitutions that retain one of more of the biological properties associated with the full-length HIO105.1 polypeptide. In some cases, variants are generated that change the post-translational processing of a HIO105.1 polypeptide. For instance, variants may have altered protein transport or protein localization characteristics or altered protein half-life compared to the native polypeptide.

As used herein, the term "HIO105.1 nucleic acid" encompasses nucleic acids with the sequence provided in or complementary to the sequence provided in SEQ ID NO:1, as well as functionally active fragments, derivatives, or orthologs thereof. A HIO105.1 nucleic acid of this invention may be DNA, derived from genomic DNA or cDNA, or RNA.

In one embodiment, a functionally active HIO105.1 nucleic acid encodes or is complementary to a nucleic acid that encodes a functionally active HIO105.1 polypeptide. Included within this definition is genomic DNA that serves as a template for a primary RNA transcript (i.e., an mRNA precursor) that requires processing, such as splicing, before encoding the functionally active HIO105.1 polypeptide. A HIO105.1 nucleic acid can include other non-coding sequences, which may or may not be transcribed; such sequences include 5' and 3' UTRs, polyadenylation signals and regulatory sequences that control gene expression, among others, as are known in the art. Some polypeptides require processing events, such as proteolytic cleavage, covalent modification, etc., in order to become fully active. Accordingly, functionally active nucleic acids may encode the mature or the pre-processed HIO105.1 polypeptide, or an intermediate form. A HIO105.1 polynucleotide can also include heterologous coding sequences, for example, sequences that encode a marker included to facilitate the purification of the fused polypeptide, or a transformation marker.

In another embodiment, a functionally active HIO105.1 nucleic acid is capable of being used in the generation of loss-of-function HIO105.1 phenotypes, for instance, via antisense suppression, co-suppression, etc.

In one preferred embodiment, a HIO105.1 nucleic acid used in the methods of this invention comprises a nucleic acid sequence that encodes or is complementary to a sequence that encodes a HIO105.1 polypeptide having at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95% or more sequence identity to the polypeptide sequence presented in SEQ ID NO:2.

In another embodiment a HIO105.1 polypeptide of the invention comprises a polypeptide sequence with at least 50% or 60% identity to the HIO105.1 polypeptide sequence of SEQ ID NO:2, and may have at least 70%, 80%, 85%, 90% or 95% or more sequence identity to the HIO105.1 polypeptide sequence of SEQ ID NO:2. In another embodiment, a HIO105.1 polypeptide comprises a polypeptide sequence with at least 50%, 60%, 70%, 80%, 85%, 90% or 95% or more sequence identity to a functionally active fragment of the polypeptide presented in SEQ ID NO:2. In yet another embodiment, a HIO105.1 polypeptide comprises a polypeptide sequence with at least 50%, 60 %, 70%, 80%, or 90% identity to the polypeptide sequence of SEQ ID NO:2 over its entire length.

In another aspect, a HIO105.1 polynucleotide sequence is at least 50% to 60% identical over its entire length to the HIO105.1 nucleic acid sequence presented as SEQ ID NO:1, or nucleic acid sequences that are complementary to such a HIO105.1 sequence,

and may comprise at least 70%, 80%, 85%, 90% or 95% or more sequence identity to the HIO105.1 sequence presented as SEQ ID NO:1 or a functionally active fragment thereof, or complementary sequences.

As used herein, "percent (%) sequence identity" with respect to a specified subject sequence, or a specified portion thereof, is defined as the percentage of nucleotides or amino acids in the candidate derivative sequence identical with the nucleotides or amino acids in the subject sequence (or specified portion thereof), after aligning the sequences and introducing gaps, if necessary to achieve the maximum percent sequence identity, as generated by the program WU-BLAST-2.0a19 (Altschul *et al.*, J. Mol. Biol. (1997) 215:403-410; website at blast.wustl.edu/blast/README.html) with search parameters set to default values. The HSP S and HSP S2 parameters are dynamic values and are established by the program itself depending upon the composition of the particular sequence and composition of the particular database against which the sequence of interest is being searched. A "% identity value" is determined by the number of matching identical nucleotides or amino acids divided by the sequence length for which the percent identity is being reported. "Percent (%) amino acid sequence similarity" is determined by doing the same calculation as for determining % amino acid sequence identity, but including conservative amino acid substitutions in addition to identical amino acids in the computation. A conservative amino acid substitution is one in which an amino acid is substituted for another amino acid having similar properties such that the folding or activity of the protein is not significantly affected. Aromatic amino acids that can be substituted for each other are phenylalanine, tryptophan, and tyrosine; interchangeable hydrophobic amino acids are leucine, isoleucine, methionine, and valine; interchangeable polar amino acids are glutamine and asparagine; interchangeable basic amino acids are arginine, lysine and histidine; interchangeable acidic amino acids are aspartic acid and glutamic acid; and interchangeable small amino acids are alanine, serine, threonine, cysteine and glycine.

Derivative nucleic acid molecules of the subject nucleic acid molecules include sequences that selectively hybridize to the nucleic acid sequence of SEQ ID NO:1. The stringency of hybridization can be controlled by temperature, ionic strength, pH, and the presence of denaturing agents such as formamide during hybridization and washing. Conditions routinely used are well known (see, *e.g.*, Current Protocol in Molecular Biology, Vol. 1, Chap. 2.10, John Wiley & Sons, Publishers (1994); Sambrook *et al.*, Molecular Cloning, Cold Spring Harbor (1989)). In some embodiments, a nucleic acid

molecule of the invention is capable of hybridizing to a nucleic acid molecule containing the nucleotide sequence of SEQ ID NO:1 under stringent hybridization conditions that are: prehybridization of filters containing nucleic acid for 8 hours to overnight at 65° C in a solution comprising 6X single strength citrate (SSC) (1X SSC is 0.15 M NaCl, 0.015 M Na citrate; pH 7.0), 5X Denhardt's solution, 0.05% sodium pyrophosphate and 100 μg/ml herring sperm DNA; hybridization for 18-20 hours at 65° C in a solution containing 6X SSC, 1X Denhardt's solution, 100 μg/ml yeast tRNA and 0.05% sodium pyrophosphate; and washing of filters at 65° C for 1 h in a solution containing 0.1X SSC and 0.1% SDS (sodium dodecyl sulfate). In other embodiments, moderately stringent hybridization conditions are used that are: pretreatment of filters containing nucleic acid for 6 h at 40° C in a solution containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.1% PVP, 0.1% Ficoll, 1% BSA, and 500 μg/ml denatured salmon sperm DNA; hybridization for 18-20 h at 40° C in a solution containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 μg/ml salmon sperm DNA, and 10% (wt/vol) dextran sulfate; followed by washing twice for 1 hour at 55° C in a solution containing 2X SSC and 0.1% SDS. Alternatively, low stringency conditions can be used that comprise: incubation for 8 hours to overnight at 37° C in a solution comprising 20% formamide, 5 x SSC, 50 mM sodium phosphate (pH 7.6), 5X Denhardt's solution, 10% dextran sulfate, and 20 μg/ml denatured sheared salmon sperm DNA; hybridization in the same buffer for 18 to 20 hours; and washing of filters in 1 x SSC at about 37° C for 1 hour.

As a result of the degeneracy of the genetic code, a number of polynucleotide sequences encoding a HIO105.1 polypeptide can be produced. For example, codons may be selected to increase the rate at which expression of the polypeptide occurs in a particular host species, in accordance with the optimum codon usage dictated by the particular host organism (see, e.g., Nakamura *et al.*, 1999). Such sequence variants may be used in the methods of this invention.

The methods of the invention may use orthologs of the *Arabidopsis* HIO105.1. Methods of identifying the orthologs in other plant species are known in the art. Normally, orthologs in different species retain the same function, due to presence of one or more protein motifs and/or 3-dimensional structures. In evolution, when a gene duplication event follows speciation, a single gene in one species, such as *Arabidopsis*, may correspond to multiple genes (paralogs) in another. As used herein, the term "orthologs" encompasses paralogs. When sequence data is available for a particular plant

species, orthologs are generally identified by sequence homology analysis, such as BLAST analysis, usually using protein bait sequences. Sequences are assigned as a potential ortholog if the best hit sequence from the forward BLAST result retrieves the original query sequence in the reverse BLAST (Huynen MA and Bork P, Proc Natl Acad Sci (1998) 95:5849-5856; Huynen MA *et al.*, Genome Research (2000) 10:1204-1210). Programs for multiple sequence alignment, such as CLUSTAL (Thompson JD *et al.*, 1994, Nucleic Acids Res 22:4673-4680) may be used to highlight conserved regions and/or residues of orthologous proteins and to generate phylogenetic trees. In a phylogenetic tree representing multiple homologous sequences from diverse species (e.g., retrieved through BLAST analysis), orthologous sequences from two species generally appear closest on the tree with respect to all other sequences from these two species. Structural threading or other analysis of protein folding (e.g., using software by ProCeryon, Biosciences, Salzburg, Austria) may also identify potential orthologs. Nucleic acid hybridization methods may also be used to find orthologous genes and are preferred when sequence data are not available. Degenerate PCR and screening of cDNA or genomic DNA libraries are common methods for finding related gene sequences and are well known in the art (see, e.g., Sambrook, 1989; Dieffenbach and Dveksler, 1995). For instance, methods for generating a cDNA library from the plant species of interest and probing the library with partially homologous gene probes are described in Sambrook *et al.* A highly conserved portion of the *Arabidopsis* HIO105.1 coding sequence may be used as a probe. HIO105.1 ortholog nucleic acids may hybridize to the nucleic acid of SEQ ID NO:1 under high, moderate, or low stringency conditions. After amplification or isolation of a segment of a putative ortholog, that segment may be cloned and sequenced by standard techniques and utilized as a probe to isolate a complete cDNA or genomic clone. Alternatively, it is possible to initiate an EST project to generate a database of sequence information for the plant species of interest. In another approach, antibodies that specifically bind known HIO105.1 polypeptides are used for ortholog isolation (see, e.g., Harlow and Lane, 1988, 1999). Western blot analysis can determine that a HIO105.1 ortholog (i.e., an orthologous protein) is present in a crude extract of a particular plant species. When reactivity is observed, the sequence encoding the candidate ortholog may be isolated by screening expression libraries representing the particular plant species. Expression libraries can be constructed in a variety of commercially available vectors, including lambda gt11, as described in Sambrook, *et al.*, 1989. Once the candidate ortholog(s) are identified by any of these means, candidate orthologous sequence are used as bait (the "query") for the

reverse BLAST against sequences from *Arabidopsis* or other species in which HIO105.1 nucleic acid and/or polypeptide sequences have been identified.

HIO105.1 nucleic acids and polypeptides may be obtained using any available method. For instance, techniques for isolating cDNA or genomic DNA sequences of interest by screening DNA libraries or by using polymerase chain reaction (PCR), as previously described, are well known in the art. Alternatively, nucleic acid sequence may be synthesized. Any known method, such as site directed mutagenesis (Kunkel TA *et al.*, 1991), may be used to introduce desired changes into a cloned nucleic acid.

In general, the methods of the invention involve incorporating the desired form of the HIO105.1 nucleic acid into a plant expression vector for transformation of in plant cells, and the HIO105.1 polypeptide is expressed in the host plant.

An isolated HIO105.1 nucleic acid molecule is other than in the form or setting in which it is found in nature and is identified and separated from least one contaminant nucleic acid molecule with which it is ordinarily associated in the natural source of the HIO105.1 nucleic acid. However, an isolated HIO105.1 nucleic acid molecule includes HIO105.1 nucleic acid molecules contained in cells that ordinarily express HIO105.1 where, for example, the nucleic acid molecule is in a chromosomal location different from that of natural cells.

Generation of Genetically Modified Plants with an Altered Oil Content Phenotype

HIO105.1 nucleic acids and polypeptides may be used in the generation of genetically modified plants having a modified oil content phenotype. As used herein, a "modified oil content phenotype" may refer to modified oil content in any part of the plant; the modified oil content is often observed in seeds. In a preferred embodiment, altered expression of the HIO105.1 gene in a plant is used to generate plants with a high oil phenotype.

The methods described herein are generally applicable to all plants. Although activation tagging and gene identification is carried out in *Arabidopsis*, the HIO105.1 gene (or an ortholog, variant or fragment thereof) may be expressed in any type of plant. In a preferred embodiment, the invention is directed to oil-producing plants, which produce and store triacylglycerol in specific organs, primarily in seeds. Such species include soybean (*Glycine max*), rapeseed and canola (including *Brassica napus*, *B. campestris*), sunflower (*Helianthus annus*), cotton (*Gossypium hirsutum*), corn (*Zea mays*), cocoa (*Theobroma cacao*), safflower (*Carthamus tinctorius*), oil palm (*Elaeis guineensis*),

coconut palm (*Cocos nucifera*), flax (*Linum usitatissimum*), castor (*Ricinus communis*) and peanut (*Arachis hypogaea*). The invention may also be directed to fruit- and vegetable-bearing plants, grain-producing plants, nut-producing plants, rapid cycling *Brassica* species, alfalfa (Medicago sativa), tobacco (Nicotiana), turfgrass (Poaceae family), other forage crops, and wild species that may be a source of unique fatty acids.

The skilled artisan will recognize that a wide variety of transformation techniques exist in the art, and new techniques are continually becoming available. Any technique that is suitable for the target host plant can be employed within the scope of the present invention. For example, the constructs can be introduced in a variety of forms including, but not limited to as a strand of DNA, in a plasmid, or in an artificial chromosome. The introduction of the constructs into the target plant cells can be accomplished by a variety of techniques, including, but not limited to *Agrobacterium*-mediated transformation, electroporation, microinjection, microprojectile bombardment calcium-phosphate-DNA co-precipitation or liposome-mediated transformation of a heterologous nucleic acid. The transformation of the plant is preferably permanent, *i.e.* by integration of the introduced expression constructs into the host plant genome, so that the introduced constructs are passed onto successive plant generations. Depending upon the intended use, a heterologous nucleic acid construct comprising an HIO105.1 polynucleotide may encode the entire protein or a biologically active portion thereof.

In one embodiment, binary Ti-based vector systems may be used to transfer polynucleotides. Standard *Agrobacterium* binary vectors are known to those of skill in the art, and many are commercially available (e.g., pBI121 Clontech Laboratories, Palo Alto, CA).

The optimal procedure for transformation of plants with *Agrobacterium* vectors will vary with the type of plant being transformed. Exemplary methods for *Agrobacterium*-mediated transformation include transformation of explants of hypocotyl, shoot tip, stem or leaf tissue, derived from sterile seedlings and/or plantlets. Such transformed plants may be reproduced sexually, or by cell or tissue culture. *Agrobacterium* transformation has been previously described for a large number of different types of plants and methods for such transformation may be found in the scientific literature. Of particular relevance are methods to transform commercially important crops, such as rapeseed (De Block *et al.*, 1989), sunflower (Everett *et al.*, 1987), and soybean (Christou *et al.*, 1989; Kline *et al.*, 1987).

Expression (including transcription and translation) of HIO105.1 may be regulated with respect to the level of expression, the tissue type(s) where expression takes place and/or developmental stage of expression. A number of heterologous regulatory sequences (e.g., promoters and enhancers) are available for controlling the expression of a HIO105.1 nucleic acid. These include constitutive, inducible and regulatable promoters, as well as promoters and enhancers that control expression in a tissue- or temporal-specific manner. Exemplary constitutive promoters include the raspberry E4 promoter (U.S. Patent Nos. 5,783,393 and 5,783,394), the 35S CaMV (Jones JD *et al*, 1992), the CsVMV promoter (Verdaguer B *et al.*, 1998) and the melon actin promoter (published PCT application WO0056863). Exemplary tissue-specific promoters include the tomato E4 and E8 promoters (U.S. Patent No. 5,859,330) and the tomato 2AII gene promoter (Van Haaren MJJ *et al.*, 1993).

In one preferred embodiment, HIO105.1 expression is under control of regulatory sequences from genes whose expression is associated with early seed and/or embryo development. Legume genes whose promoters are associated with early seed and embryo development include *V. faba legumin* (Baumlein *et al.*, 1991, Mol Gen Genet 225:121-8; Baumlein *et al.*, 1992, Plant J 2:233-9), *V. faba usp* (Fiedler *et al.*, 1993, Plant Mol Biol 22:669-79), pea *convicilin* (Bown *et al.*, 1988, Biochem J 251:717-26), pea *lectin* (dePater *et al.*, 1993, Plant Cell 5:877-86), *P. vulgaris beta phaseolin* (Bustos *et al.*, 1991, EMBO J 10:1469-79), *P. vulgaris DLEC2* and *PHS* [beta] (Bobb *et al.*, 1997, Nucleic Acids Res 25:641-7), and soybean *beta-Conglycinin*, 7S storage protein (Chamberland *et al.*, 1992, Plant Mol Biol 19:937-49). Cereal genes whose promoters are associated with early seed and embryo development include rice *glutelin* ("GluA-3," Yoshihara and Takaiwa, 1996, Plant Cell Physiol 37:107-11; "GluB-1," Takaiwa *et al.*, 1996, Plant Mol Biol 30:1207-21; Washida *et al.*, 1999, Plant Mol Biol 40:1-12; "Gt3," Leisy *et al.*, 1990, Plant Mol Biol 14:41-50), rice *prolamin* (Zhou & Fan, 1993, Transgenic Res 2:141-6), wheat *prolamin* (Hammond-Kosack *et al.*, 1993, EMBO J 12:545-54), maize *zein* (Z4, Matzke *et al.*, 1990, Plant Mol Biol 14:323-32), and barley *B-hordeins* (Entwistle *et al.*, 1991, Plant Mol Biol 17:1217-31). Other genes whose promoters are associated with early seed and embryo development include oil palm GLO7A (7S globulin, Morcillo *et al.*, 2001, Physiol Plant 112:233-243), *Brassica napus napin*, 2S storage protein, and napA gene (Josefsson *et al.*, 1987, J Biol Chem 262:12196-201; Stalberg *et al.*, 1993, Plant Mol Biol 1993 23:671-83; Ellerstrom *et al.*, 1996, Plant Mol Biol 32:1019-27), *Brassica napus oleosin* (Keddie *et al.*, 1994, Plant Mol Biol 24:327-40), *Arabidopsis oleosin* (Plant *et al.*, 1994, Plant Mol Biol

25:193-205), *Arabidopsis* FAE1 (Rossak *et al.*, 2001, Plant Mol Biol 46:717-25), *Canavalia gladiata* conA (Yamamoto *et al.*, 1995, Plant Mol Biol 27:729-41), and *Catharanthus roseus* strictosidine synthase (Str, Ouwerkerk and Memelink, 1999, Mol Gen Genet 261:635-43). In another preferred embodiment, regulatory sequences from genes expressed during oil biosynthesis are used (see, e.g., US Pat No: 5,952, 544). Alternative promoters are from plant storage protein genes (Bevan *et al.*, 1993, Philos Trans R Soc Lond B Biol Sci 342:209-15).

In yet another aspect, in some cases it may be desirable to inhibit the expression of endogenous HIO105.1 in a host cell. Exemplary methods for practicing this aspect of the invention include, but are not limited to antisense suppression (Smith, *et al.*,1988; van der Krol *et al.*, 1988); co-suppression (Napoli, *et al.*, 1990); ribozymes (PCT Publication WO 97/10328); and combinations of sense and antisense (Waterhouse, *et al.*, 1998). Methods for the suppression of endogenous sequences in a host cell typically employ the transcription or transcription and translation of at least a portion of the sequence to be suppressed. Such sequences may be homologous to coding as well as non-coding regions of the endogenous sequence. Antisense inhibition may use the entire cDNA sequence (Sheehy *et al.*, 1988), a partial cDNA sequence including fragments of 5' coding sequence, (Cannon *et al.*, 1990), or 3' non-coding sequences (Ch'ng *et al.*, 1989). Cosuppression techniques may use the entire cDNA sequence (Napoli *et al.*, 1990; van der Krol *et al.*, 1990), or a partial cDNA sequence (Smith *et al.*, 1990).

Standard molecular and genetic tests may be performed to further analyze the association between a gene and an observed phenotype. Exemplary techniques are described below.

1. DNA/RNA analysis

The stage- and tissue-specific gene expression patterns in mutant versus wild-type lines may be determined, for instance, by in situ hybridization. Analysis of the methylation status of the gene, especially flanking regulatory regions, may be performed. Other suitable techniques include overexpression, ectopic expression, expression in other plant species and gene knock-out (reverse genetics, targeted knock-out, viral induced gene silencing [VIGS, see Baulcombe D, 1999]).

In a preferred application expression profiling, generally by microarray analysis, is used to simultaneously measure differences or induced changes in the expression of many different genes. Techniques for microarray analysis are well known in the art (Schena M

et al., Science (1995) 270:467-470; Baldwin D *et al.*, 1999; Dangond F, Physiol Genomics (2000) 2:53-58; van Hal NL *et al.*, J Biotechnol (2000) 78:271-280; Richmond T and Somerville S, Curr Opin Plant Biol (2000) 3:108-116). Expression profiling of individual tagged lines may be performed. Such analysis can identify other genes that are coordinately regulated as a consequence of the overexpression of the gene of interest, which may help to place an unknown gene in a particular pathway.

2. Gene Product Analysis

Analysis of gene products may include recombinant protein expression, antisera production, immunolocalization, biochemical assays for catalytic or other activity, analysis of phosphorylation status, and analysis of interaction with other proteins via yeast two-hybrid assays.

3. Pathway Analysis

Pathway analysis may include placing a gene or gene product within a particular biochemical, metabolic or signaling pathway based on its mis-expression phenotype or by sequence homology with related genes. Alternatively, analysis may comprise genetic crosses with wild-type lines and other mutant lines (creating double mutants) to order the gene in a pathway, or determining the effect of a mutation on expression of downstream "reporter" genes in a pathway.

Generation of Mutated Plants with an Altered Oil Content Phenotype

The invention further provides a method of identifying plants that have mutations in endogenous HI0105.1 that confer altered oil content, and generating altered oil content progeny of these plants that are not genetically modified. In one method, called "TILLING" (for **t**argeting **i**nduced **l**ocal **l**esions **in** **g**enomes), mutations are induced in the seed of a plant of interest, for example, using EMS treatment. The resulting plants are grown and self-fertilized, and the progeny are used to prepare DNA samples. HI0105.1-specific PCR is used to identify whether a mutated plant has a HI0105.1 mutation. Plants having HI0105.1 mutations may then be tested for altered oil content, or alternatively, plants may be tested for altered oil content, and then HI0105.1-specific PCR is used to determine whether a plant having altered oil content has a mutated HI0105.1 gene. TILLING can identify mutations that may alter the expression of specific genes or the activity of proteins encoded by these genes (see Colbert et al (2001) Plant Physiol 126:480-484; McCallum et al (2000) Nature Biotechnology 18:455-457).

In another method, a candidate gene/Quantitative Trait Locus (QTLs) approach can be used in a marker-assisted breeding program to identify alleles of or mutations in the HI0105.1 gene or orthologs of HI0105.1 that may confer altered oil content (see Bert et al., Theor Appl Genet. 2003 Jun;107(1):181-9; and Lionneton et al, Genome. 2002 Dec;45(6):1203-15). Thus, in a further aspect of the invention, a HI0105.1 nucleic acid is used to identify whether a plant having altered oil content has a mutation in endogenous HI0105.1 or has a particular allele that causes altered oil content.

While the invention has been described with reference to specific methods and embodiments, it will be appreciated that various modifications and changes may be made without departing from the invention. All publications cited herein are expressly incorporated herein by reference for the purpose of describing and disclosing compositions and methodologies that might be used in connection with the invention. All cited patents, patent applications, and sequence information in referenced websites and public databases are also incorporated by reference.

EXAMPLES

EXAMPLE 1

Generation of Plants with a HIO105.1 Phenotype by Transformation with an Activation Tagging Construct

Mutants were generated using the activation tagging "ACTTAG" vector, pSKI015 (GI#6537289; Weigel D *et al.*, 2000). Standard methods were used for the generation of *Arabidopsis* transgenic plants, and were essentially as described in published application PCT WO0183697. Briefly, T0 *Arabidopsis* (Col-0) plants were transformed with *Agrobacterium* carrying the pSKI015 vector, which comprises T-DNA derived from the *Agrobacterium* Ti plasmid, an herbicide resistance selectable marker gene, and the 4X CaMV 35S enhancer element. Transgenic plants were selected at the T1 generation based on herbicide resistance.

T3 seed pools were analyzed by Near Infrared Spectroscopy (NIR) intact at time of harvest. NIR infrared spectra were captured using a Bruker 22 N/F. Bruker Software was used to estimate total seed oil and total seed protein content using data from NIR analysis and reference methods according to the manufacturers instructions. Oil contents predicted by our calibration (PDX Oil 3, Predicts Hexane Extracted Oil) were compared for 40,000 individual ACTTAG lines. To identify high oil lines the NIR oil result was compared to the mean oil result for all ACTTAG lines planted on the same day (Relative oil content).

Subsequent to seed compositional analysis ACTTAG flanking sequence was determined. 22,000 lines with recovered flanking sequences were considered in this analysis. From 22,000 placed ACTTAG lines 819 (~4%) had high oil (defined as oil content of >=107% of the planting day average oil content). The genome coordinates in this subset of high oil ACTTAG lines was evaluated and lines having 2 or more independent ACTTAG insertions, also displaying high oil, were identified and the flanking sequence confirmed by PCR and sequencing.

Line W000139144 (IN045421) had a NIR determined oil content of 41.0% relative to a planting day average oil content of 37.1% (110% of PDA). Line W000147756 (IN051997) had a NIR determined oil content of 37.1% relative to a planting day average oil content of 34.6%(107% of PDA).

Line W000139144 (IN045421) had a confirmed ACTTAG insertion on Chromosome 2 at bp 16680674.

Line W000147756 (IN051997) had a confirmed ACTTAG insertion on Chromosome 2 at bp 16680496.

Based on the presence of an enhancer-containing T-DNA insert near this candidate gene in more than one independent mutant line showing a high oil phenotype, we concluded that the insert was linked with the trait. The actual insert number for these mutant lines was unknown.

EXAMPLE 2

Characterization of the T-DNA Insertion in Plants Exhibiting the Altered Oil Content Phenotype.

We performed standard molecular analyses, essentially as described in patent application PCT WO0183697, to determine the site of the T-DNA insertion associated with the altered oil content phenotype. Briefly, genomic DNA was extracted from plants exhibiting the altered oil content phenotype. PCR, using primers specific to the pSKI015 vector, confirmed the presence of the 35S enhancer in plants from lines IN045421 and IN051997, and Southern blot analysis verified the genomic integration of the ACTTAG T-DNA and showed the presence of the T-DNA insertions in each of the transgenic lines.

Inverse PCR was used to recover genomic DNA flanking the T-DNA insertion, which was then subjected to sequence analysis using a basic BLASTN search and/or a search of the *Arabidopsis* Information Resource (TAIR) database (available at the arabidopsis.org website). For ACTTAG line IN045421, there was sequence identity to nucleotides 100702-100973 on Arabidopsis clone T28M21 map CIC10A06F22O13 chromosome 2 (GI#20198304), placing the left border junction upstream from 100973 (GI#20198304). The opposite flank (predicted right border junction) of this insert was not determined. Left border of IN045421 T-DNA was ~ 11655 bp 3' of the translation start site.

For ACTTAG line IN051997, there was sequence identity to nucleotides 100613-100784 on Arabidopsis clone T28M21 map CIC10A06F22O13 chromosome 2 (GI#20198304), placing the left border junction downstream from 100613 (GI#20198304). The opposite flank (predicted right border junction) of this insert was not determined. Left border IN051997 T-DNA was ~ 11477 bp 3' of the translation start site.

EXAMPLE 3

Analysis of *Arabidopsis* HIO105.1 Sequence

Sequence analyses were performed with BLAST (Altschul *et al.*, 1997, J. Mol. Biol. 215:403-410), PFAM (Bateman *et al.*, 1999, Nucleic Acids Res 27:260-262), PSORT (Nakai K, and Horton P, 1999,Trends Biochem Sci 24:34-6), and/or CLUSTAL (Thompson JD *et al.*, 1994, Nucleic Acids Res 22:4673-4680).

TBLASTN against est_others results:

There are two Arabidopsis ESTs that matche the candidate gene At2g40040.

- gi:509192
- gi:496826

There are ESTs from diverse plant species showing similarity to At2g40040. If possible, ESTs contigs of each species were made. The top hit for each of the following species are listed below and included in the "Orthologue Table": Triticum aestivum, Gossypium hirsutum, Zea mays, Glycine max, Populus tremula, Oryza sativa, Lycopersicon esculentum, Solanum tuberosum, Brassica napus, and Hordeum vulgare.

1. One EST from wheat

>gi|17143156 WHE0460_F10_K19ZS Wheat Fusarium graminearum infected spike cDNA library Triticum aestivum cDNA clone WHE0460_F10_K19, MRNA sequence

2. One EST from cotton

>gi|5047799 BNLGHi12104 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar to (U55219) defective chloroplasts and leaves; required for chloroplast development and palisade cell differentiation in leaves [Lycopersicon esculentum] gi|1323698 (U55278) DCL [Solanum lycopersicum], MRNA sequence

3. One EST contig from maize

The contigged sequence is presented as SEQ ID NO: 3.

4. One EST from soybean

>gi|9564119 sp53d10.y1 Gm-c1043 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1043-1964 5' similar to TR:O04207 O04207 UNKNOWN PROTEIN, MRNA sequence

5. One EST contig from poplars

The contigged sequence is presented as SEQ ID NO: 4.

6. One EST from rice

>gi|13420599|dbj|AU174689.1|AU174689 AU174689 Rice cDNA from leaf Oryza sativa (japonica cultivar-group) cDNA clone F10212

7. One EST contig from tomato

The contigged sequence is presented as SEQ ID NO: 5.

8. One EST contig from potato

The contigged sequence is presented as SEQ ID NO: 6.

BLASTP against all.aa results:

The protein At2g40040 has significant homology to proteins from plant, bacteria and animal. The top 8 BLAST results for At2g40040 are listed below and are included in the "Orthologue Table.

1. At2g40040 itself (several redundant entries)

>gi|15225617|ref|NP_s81533.1| unknown protein; protein id: At2g40040.1 [Arabidopsis thaliana] >gi|25408660|pir||E84824 hypothetical protein At2g40040 [imported] - Arabidopsis thaliana >gi|2088657|gb|AAB95289.1| unknown protein [Arabidopsis thaliana]

Score = 3612 (1276.5 bits), Expect = 0., P = 0.

2. A protein similar to DNA-directed RNA polymerase subunit A' from rice

>gi|21902060|dbj|BAC05608.1| similar to DNA-directed RNA polymerase subunit A' [Oryza sativa (japonica cultivar-group)]

Score = 323 (118.8 bits), Expect = 2.9e-49, Sum P(2) = 2.9e-49

3. At3g46630 from Arabidopsis (several redundant entries)

>gi|15232596|ref|NP_190247.1| putative protein; protein id: At3g46630.1, supported by cDNA: 21432. [Arabidopsis thaliana] >gi|11357353|pir||T45600 hypothetical protein F12A12.150 - Arabidopsis thaliana >gi|6523066|emb|CAB62333.1| putative protein [Arabidopsis thaliana] >gi|21554240|gb|AAM63315.1| unknown [Arabidopsis thaliana] >gi|26452933|dbj|BAC43543.1| unknown protein [Arabidopsis thaliana]

Score = 285 (105.4 bits), Expect = 1.5e-23, P = 1.5e-23

4. DCL protein from coffee

>gi|18650724|emb|CAD12248.1| DCL protein [Coffea arabica]

Score = 258 (95.9 bits), Expect = 1.2e-20, P = 1.2e-20

5. At5g04290 from Arabidopsis

>gi|15237667|ref|NP_196049.1| glycine-rich protein; protein id: At5g04290.1 [Arabidopsis thaliana]

Score = 279 (103.3 bits), Expect = 7.3e-20, P = 7.3e-20

6. At1g45261 from Arabidopsis (several redundant entries)

>gi|15220071|ref|NP_175134.1| defective chloroplasts and leaves (DCL) protein, putative; protein id: At1g45261.1 [Arabidopsis thaliana] >gi|22330069|ref|NP_683398.1| similar to defective chloroplasts and leaves (DCL) protein, putative; protein id: At1g45230.1 [Arabidopsis thaliana] >gi|25334999|pir||C96510 hypothetical protein F2G19.1 [imported] - Arabidopsis thaliana

>gi|12321014|gb|AAG50632.1|AC083835_17 defective chloroplasts and leaves (DCL) protein, putative [Arabidopsis thaliana] >gi|22135799|gb|AAM91086.1| At1g45261/F2G19.1 [Arabidopsis thaliana]

Score = 232 (86.7 bits), Expect = 1.9e-19, Sum P(2) = 1.9e-19

7. DCL protein from tomato (several redundant entries)

>gi|6014934|sp|Q42463|DCL_LYCES DCL PROTEIN, CHLOROPLAST PRECURSOR (DEFECTIVE CHLOROPLASTS AND LEAVES PROTEIN) >gi|7459239|pir||S71749 DCL protein precursor, chloroplast - tomato >gi|1305531|gb|AAC49433.1| defective chloroplasts and leaves; required for chloroplast development and palisade cell differentiation in leaves [Lycopersicon esculentum] >gi|1323698|gb|AAC49434.1| DCL

Score = 243 (90.6 bits), Expect = 5.0e-19, P = 5.0e-19

8. At1g63020 from Arabidopsis (several redundant entries)

>gi|15221663|ref|NP_176490.1| RNA polymerase IIA large subunit, putative; protein id: At1g63020.1 [Arabidopsis thaliana]

Score = 178 (67.7 bits), Expect = 1.4e-16, Sum P(2) = 1.4e-16

The following sequences are other redundant entries of At4g18650. However, they differ from the sequences listed above by a few nucleotides. This is likely to be the result of sequencing errors or of single nucleotide polymorphisms with little or no effect on activity.

>gi|25404390|pir||D96655 hypothetical protein F16M19.19 [imported] - Arabidopsis thaliana >gi|8493592|gb|AAF75815.1|AC011000_18 Contains similarity to DNA-directed RNA Polymerase alpha-subunit from Methanococcus jannaschii gi|3915860 and contains Ribosomal protein L36 PF|00444 and RNA-Polymerase alpha-subunit PF|00623 domains. EST gb|AA597311 comes from this gene. [Arabidopsis

thaliana] >gi|12323252|gb|AAG51604.1|AC010795_8 RNA polymerase IIA largest subunit, putative; 12353-6556 [Arabidopsis thaliana]

Ortholog Gene Name	**Species**	**GI #**	**% ID to HIO105.1**	**Score(s) (BLAST, Clustal, etc.)**
A rice protein similar to DNA-directed RNA polymerase subunit A'	Oryza sativa (japonica cultivar-group)	>gi|21902060|	Length = 1507 *Identities = 96/329 (29%), Positives = 144/329 (43%)*	*BLASTP Score = 323 (118.8 bits), Expect = 2.9e-49, Sum P(2) = 2.9e-49*
DCL protein from coffee	Coffea arabica	>gi|18650724|	Length = 224 *Identities = 47/121 (38%), Positives = 74/121 (61%)*	*BLASTP Score = 258 (95.9 bits), Expect = 1.2e-20, P = 1.2e-20*
DCL protein precursor from tomato	Lycopersicon esculentum	>gi|6014934| >gi|7459239| >gi|1305531| >gi|1323698|	Length = 224 *Identities = 45/112 (40%), Positives = 69/112 (61%)*	*BLASTP Score = 243 (90.6 bits), Expect = 5.0e-19, P = 5.0e-19*
One EST from wheat	Triticum aestivum	gi|17143156	Length = 509 *Identities = 49/117 (41%), Positives = 75/117 (64%), Frame = +2*	*TBLASTN Score = 261 (96.9 bits), Expect = 4.3e-21, P = 4.3e-21*
One EST from cotton	Gossypium hirsutum	gi|5047799	Length = 604 *Identities = 53/113 (46%), Positives = 75/113 (66%), Frame = +3*	*TBLASTN Score = 288 (106.4 bits), Expect = 2.1e-25, P = 2.1e-25*
One EST contig from maize	Zea mays	gi|16921262 gi|18176077 gi|18173887	Length = 535 *Identities = 58/142 (40%), Positives = 85/142 (59%), Frame = -1*	*TBLASTN Score = 290 (107.1 bits), Expect = 8.6e-25, P = 8.6e-25*
One EST from soybean	Glycine max	gi|9564119	Length = 522 *Identities = 85/184 (46%), Positives = 118/184 (64%), Frame = +3*	*TBLASTN Score = 420 (152.9 bits), Expect = 1.1e-38, P = 1.1e-38*
One EST contig from poplars	Populus tremula	gi|24063542 gi|24067292 gi|24074960 gi|23976797 gi|24018305 gi|24076685 gi|24056443	Length = 694 *Identities = 38/98 (38%), Positives = 58/98 (59%), Frame = +1*	*TBLASTN Score = 210 (79.0 bits), Expect = 2.7e-16, P = 2.7e-16*

One EST from rice	Oryza sativa	>gi|13420599|	Length = 413 *Identities = 50/110 (45%), Positives = 75/110 (68%), Frame = +3*	*TBLASTN Score = 283 (104.7 bits), Expect = 9.4e-22, P = 9.4e-22*
One EST contig from tomato	Lycopersicon esculentum	gi|12628377 gi|14686572	Length = 1057 *Identities = 50/113 (44%), Positives = 75/113 (66%), Frame = +3*	*TBLASTN Score = 279 (103.3 bits), Expect = 1.0e-23, P = 1.0e-23*
One EST contig from potato	Solanum tuberosum	gi|11035940 gi|13616171 gi|14642630 gi|19819607 gi|21923364 gi|21923365	Length = 935 *Identities = 50/113 (44%), Positives = 75/113 (66%), Frame = +2*	*TBLASTN Score = 279 (103.3 bits), Expect = 5.9e-24, P = 5.9e-24*

Closest Arabidopsis homologs:

At3g46630	Arabidopsis thaliana	>gi|15232596| >gi|11357353| >gi|6523066| >gi|21554240| >gi|26452933|	Length = 207 *Identities = 52/112 (46%), Positives = 73/112 (65%)*	*BLASTP Score = 285 (105.4 bits), Expect = 1.5e-23, P = 1.5e-23*
At5g04290	Arabidopsis thaliana	>gi|15237667|	Length = 1493 *Identities = 141/562 (25%), Positives = 207/562 (36%)*	*BLASTP Score = 279 (103.3 bits), Expect = 7.3e-20, P = 7.3e-20*
At1g45230	Arabidopsis thaliana	>gi|15220071| >gi|22330069| >gi|25334999| >gi|12321014| >gi|22135799|	Length = 219 *Identities = 44/118 (37%), Positives = 68/118 (57%)*	*BLASTP Score = 232 (86.7 bits), Expect = 1.9e-19, Sum P(2) = 1.9e-19*
At1g63020	Arabidopsis thaliana	>gi|15221663| >gi|25404390| >gi|8493592| >gi|12323252	Length = 1432 *Identities = 37/110 (33%), Positives = 63/110 (57%)*	*BLASTP Score = 178 (67.7 bits), Expect = 1.4e-16, Sum P(2) = 1.4e-16*

At2g40040 is a non-secretory protein that lacks a transmembrane domain (predicted by TMHMM) and signal peptide (predicted by SignalP). At2g40040 is likely to be localized to the cytoplasm or to the nucleus (44% cytoplasmic, 36% nuclear by Psort2).

Pfam analysis showed that At2g40040 has no known protein motifs. BLAST analysis described above suggests that At2g40040 shares homology to a small number of plant proteins.

EXAMPLE 4

Generation of Mutated Plants with a HIO105.1 Phenotype

The invention further provides a method of identifying plants that have mutations in, or an allele of, endogenous HIO105.1 that confer a HIO105.1 phenotype, and generating progeny of these plants that also have the HIO105.1 phenotype and are not genetically modified. In one method, called "TILLING" (for targeting induced local lesions in genomes), mutations are induced in the seed of a plant of interest, for example, using EMS treatment. The resulting plants are grown and self-fertilized, and the progeny are used to prepare DNA samples. HIO105.1-specific PCR is used to identify whether a mutated plant has a HIO105.1 mutation. Plants having HIO105.1 mutations may then be tested for the HIO105.1 phenotype, or alternatively, plants may be tested for the HIO105.1 phenotype, and then HIO105.1-specific PCR is used to determine whether a plant having the HIO105.1 phenotype has a mutated HIO105.1 gene. TILLING can identify mutations that may alter the expression of specific genes or the activity of proteins encoded by these genes (see Colbert *et al.* (2001) Plant Physiol 126:480-484; McCallum *et al.* (2000) Nature Biotechnology 18:455-457).

In another method, a candidate gene/Quantitative Trait Locus (QTLs) approach can be used in a marker-assisted breeding program to identify alleles of or mutations in the HIO105.1 gene or orthologs of HIO105.1 that may confer the HIO105.1 phenotype (see Foolad *et al.*, Theor Appl Genet. (2002) 104(6-7):945-958; Rothan *et al.*, Theor Appl Genet (2002) 105(1):145-159); Dekkers and Hospital, Nat Rev Genet. (2002) Jan;3(1):22-32).

Thus, in a further aspect of the invention, a HIO105.1 nucleic acid is used to identify whether a plant having a HIO105.1 phenotype has a mutation in endogenous HIO105.1 or has a particular allele that causes the HIO105.1 phenotype compared to plants lacking the mutation or allele, and generating progeny of the identified plant that have inherited the HIO105.1 mutation or allele and have the HIO105.1 phenotype.

REFERENCES

Adams, M.D. *et al.*, Science 287(5461):2195, 2000.

Ageorges, A, *et al.*, Plant Physiol. Biochem. 38:177-185, 2000.

Agulnik, A.I., *et al.*, Genome 10(9):926-929, 1999.

Altschul, S.F. *et al.*, J. Mol. Biol. 215:403-410, 1990.

Altschul, S.F. *et al.*, Nucleic Acids Res. 25:3389-3402, 1997.

Asamizu, E. *et al.*, DNA Res. 5:379-391, 1998.

Ausubel FM *et al.*, Current Protocols in Molecular Biology, John Wiley & Sons, New York, N.Y., 1993.

Baldwin D *et al.*, Cur Opin Plant Biol. 2(2):96-103, 1999.

Bateman *et al.*, 1999, Nucleic Acids Res 27:260-262 (website at pfam.wustl.edu).

Baulcombe D, Arch Virol Suppl 15:189-201, 1999.

Cannon *et al.*, Plant Molec. Biol. (1990) 15:39-47.

Ch'ng *et al.*, Proc. Natl. Acad. Sci. USA (1989) 86:10006-10010

Christensen S *et al.*, 9th International Conference on *Arabidopsis* Research. Univ. of Wisconsin-Madison, June 24-28, 1998. Abstract 165.

Christou *et al.*, Proc. Natl. Acad. Sci USA (1989) 86:7500-7504.

Cough, SJ and Bent, AF, the Plant Journal 16(6): 735-743, 1998.

Davies, C., *et al.*, Plant Science 147:93-100, 1999.

De Block *et al.*, Plant Physiol. (1989) 91:694-701.

Defeo-Jones, D., *et al.*, Nature 352(6332):251-251, 1991.

Dieffenbach C and Dveksler G (Eds.) PCR Primer: A Laboratory Manual, Cold Spring Harbor Laboratory Press, NY, 1989.

Everett *et al.*, Bio/Technology (1987) 5:1201

The FANTOM Consortium and the RIKEN Genome Exploratio Research Group Phas I & II Team, Nature 420:563-573, 2002.

Feldmann *et al.*, Science 243: 1351-1354, 1989.

Focks N and Benning C, Plant Physiol 118:91-101, 1998.

Fridborg I *et al.*, Plant Cell 11: 1019-1032, 1999.

Fujita, T *et al.*, Plant J. 5:645-654, 1994.

Geest AH and Hall TC, Plant Mol Biol 32(4):579-88, 1996.

Gelvin, S. B., Schilperoort, R. A., Varma, D. P. S., eds. Plant Molecular Biology Manual 1990.

Gibbons, R.J. *et al.*, Nat. Genet. 24(4):368-371, 2000.

Glick, BR and Thompson, JE, Eds. Methods in Plant Molecular Biology and Biotechnology, p. 213-221, CRC Press, 1993.

Harlow E and Lane D, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1988, New York.

Harlow E and Lane D, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1999, New York

Hayashi H *et al.*, Science 258: 1350-1353, 1992.

Heidelberg, J.F.*et al.*, Nature 406(6795):477-483, 2000.

Jako *et al.*, Plant Physiology 126(2):861-74, 2001.

James DW and Dooner HK (1990) Theor Appl Genet 80, 241-245.

Jensen, L.G., *et al.*, Proc. Natl. Acad. Sci. USA 93:3487-3491, 1996.

Jones JD *et al.*, Transgenic Res 1:285-297 1992.

Kaneko, T. *et al.*, DNA Res. 7:217-221, 2000.

Kardailsky I *et al.*, Science 286: 1962-1965, 1999.

Kataoka, M. *et al.*, Microgiol. Lett. 151(2):245-248, 1997.

Katavic V *et al.*, Plant Physiology 108(1):399-409, 1995.

Kawai, J. *et al.*, Nature 409(6821):685-690, 2001.

Keddie, J.S. *et al.*, EMBO J. 15:4208-4217, 1996.

Kielkopf, C.L. and Burley, S.K., Biochemistry 41(39):11711-11720, 2002.

Kline *et al.*, Nature (1987) 327:70.

Knop, C., *et al.*, Planta 213:80-91, 2001.

Komjanc, M. *et al.*, Plant Mol. Biol. 40:945-957, 1999.

Kunkel TA *et al.*, Methods Enzymol. 204:125-39, 1991.

Le Douarin, B., *et al.*, EMBO J. 15(23):6701-6715, 1996.

Lemieux B., *et al.*, 1990, Theor Appl Genet 80, 234-240.

Lin, X. *et al.*, Nature 402(6763):761-768, 1999.

Liu, J.Q. *et al.*, J. Bacteriol. 179(11):3555-3560, 1997.

Ludwig, A., *et al.*, Plant J. 24:503-509, 2000.

Makino, K. *et al.*, Lancet 361(9359):743-749, 2003.

Mao, S., *et al.*, Oncogene 14(13):1531-1539, 1997.

Meyer, S., *et al.*, Plant J. 24:869, 2000.

Muris, D.F. *et al.*, J. Cell Sci. 109(Pt. 1):73-81, 1996.

Nagase, T. *et al.*, DNA Res. 7(1):65-73, 2000.

Nakamura Y *et al.*, 1999, Nucleic Acids Res 27:292.

Napoli, *et al.*, Plant Cell 2:279-289, 1990.

Nelson, K.E.*et al.*, Nature 399(6734):323-329, 1999.

Noiraud, N., *et al.*, Plant Physiol. 122:1447-1455, 2000.

Okuley *et al.*, Plant Cell 6(1):147-158, 1994.

Omirulleh *et al.*, Plant Mol Biol. 21(3): 415-28, 1993.

Picketts, D.J. *et al.*, Mamm. Genome 9(5):400-403, 1998.

Sambrook *et al.*, Molecular Cloning: A Laboratory Manual (Second Edition), Cold

Sasaki, T., Nature 420(6913):312-316, 2002.

Sauer, N. and Stolz, J., Plant J. 6:67-77, 1994.

Spring Harbor Press, Plainview, N.Y., 1989.

Schaffer R, *et al.*, Cell 93: 1219-1229, 1998.

Sheehy *et al.*, Proc. Natl. Acad. Sci. USA (1988) 85:8805-8809.

Smith, *et al.*, Nature 334:724-726, 1988.

Smith *et al.*, Mol. Gen. Genetics (1990) 224:477-481.

Tabata, T. *et al.*, EMBO J. 10:1459-1467, 1991.

Tanaka, K. *et al.*, J. Biol Chem 275(34):26316-26321, 2000.

Theologis, A. *et al.*, Nature 408:816-820, 2000.

Thompson JD *et al.*, Nucleic Acids Res 22:4673-4680, 1994.

Tsuchiya, K.D. and Willard, H.F., Genome 11(10):849-854, 2000.

van der Krol *et al.*, Biotechniques (1988) 6:958-976.

van der Krol *et al.*, The Plant Cell (1990) 2:291-299.

Van Haaren MJJ *et al.*, Plant Mol Bio 21:625-640, 1993.

Verdaguer B *et al.*, Plant Mol Biol 37:1055-1067, 1998.

Villard, L. *et al.*, Genomics 43(2):149-155, 1997.

Walker, J.C. Plant J. 3(3):45-456, 1993.

Waterhouse, *et al.*, Proc. Natl. Acad. Sci. USA 95:13959-13964, 1998.

Weschke W., *et al.*, Plant J. 21:455-467, 2000.

Weschke W., *et al.*, Plant J. 33:395-411, 2003.

Weigel D, *et al.*, Plant Physiology, 122:1003-1013, 2000.

Weise A, *et al.*, Plant Cell 12:1345-1355, 2000.

Wilson K *et al.*, Plant Cell 8: 659-671, 1996.

Wu, J. *et al.*, Nat. Genet. 7(4):491-496, 1994.

Yadav NS *et al.*, (1993) Plant Physiol 103, 467-476.

IT IS CLAIMED:

1. A transgenic plant comprising a plant transformation vector comprising a nucleotide sequence that encodes or is complementary to a sequence that encodes a HI0105.1 polypeptide comprising the amino acid sequence of SEQ ID NO:2, or an ortholog thereof, whereby the transgenic plant has a high oil phenotype relative to control plants.

2. The transgenic plant of claim 1, which is selected from the group consisting of rapeseed, soy, corn, sunflower, cotton, cocoa, safflower, oil palm, coconut palm, flax, castor and peanut.

3. A plant part obtained from the plant according to claim 1.

4. The plant part of claim 3, which is a seed.

5. A method of producing oil comprising growing the transgenic plant of claim1 and recovering oil from said plant.

6. A method of producing a high oil phenotype in a plant, said method comprising:

a) introducing into progenitor cells of the plant a plant transformation vector comprising a nucleotide sequence that encodes or is complementary to a sequence that encodes a HI0105.1 polypeptide comprising the amino acid sequence of SEQ ID NO:2, or an ortholog thereof, and

b) growing the transformed progenitor cells to produce a transgenic plant, wherein said polynucleotide sequence is expressed, and said transgenic plant exhibits an altered oil content phenotype relative to control plants.

7. A plant obtained by a method of claim 6.

8. The plant of claim 7, which is selected from the group consisting of rapeseed, soy, corn, sunflower, cotton, cocoa, safflower, oil palm, coconut palm, flax, castor and peanut.

9. A method of generating a plant having a high oil phenotype comprising identifying a plant that has an allele in its HI0105.1 gene that results in increased oil content compared

to plants lacking the allele and generating progeny of said identified plant, wherein the generated progeny inherit the allele and have the high oil phenotype.

10. The method of claim 9 that employs candidate gene/QTL methodology.

11. The method of claim 9 that employs TILLING methodology.

SEQUENCE LISTING

<110> Agrinomics LLC

<120> GENERATION OF PLANTS WITH ALTERED OIL CONTENT

<130> AG04-007C-PC

<150> 60/464,550
<151> 2003-04-22

<160> 6

<170> PatentIn version 3.2

<210> 1
<211> 2520
<212> DNA
<213> Arabidopsis thaliana

<400> 1
atggtctcca agggagtttt gaaagaacac atcattctcc tagcaaacaa catgacttgt 60

tcagggacta tgttgggttt caattctgga ggttacaaag ctttgacccg atccttaaac 120

atcaaggctc cattcacaga agcaactctg attgcaccaa gaaaatgttt tgagaaagcg 180

gctgagaaat gtcatacaga ttcactatca acagtcgttg ggtcttgttc ttggggaaaa 240

cgcgtggatg ttggtacagg ctcacagttt gagctcctat ggaaccaaaa agagactggt 300

ctagatgata aagaagaaac agatgtgtac agcttccttc aaatggtcat atccacgaca 360

aacgcagatg catttgtttc ttcccctggt tttgatgtta cagaggaaga aatggctgaa 420

tgggctgaat cgcccgagcg agactctgct cttggagagc ccaagtttga ggacagtgcc 480

gacttccaaa atttacatga cgaaggcaag ccatcaggag ccaactggga gaaaagctct 540

tcgtgggaca atggttgcag tggtggttca gagtggggag tctcaaaaag tacaggtggt 600

gaggcaaacc cggagtctaa ttgggaaaag actacaaatg ttgagaagga agatgcttgg 660

tctagttgga acactagaaa agatgctcaa gaatcttcca agtcagattc agggggtgct 720

tggggtataa agacaaagga tgctgatgca gatacaacac caaactggga aacaagtcca 780

gctccgaagg attccattgt cccagaaaat aatgagccga cttctgacgt ttggggacac 840

aaatctgttt cagacaaatc gtgggacaaa aagaactggg gaactgaatc agctcctgct 900

gcatggggtt ctacagatgc tgctgtttgg ggttctagtg ataaaaagaa ctcggaaaca 960

gaatcagatg ctgctgcttg gggttccagg gacaaaaata attctgacgt tgggtcaggt 1020

gctggtgttt taggtccctg gaacaagaag agctcggaaa ctgagtcaaa tggtgctact 1080

tggggttcta gtgacaaaac caaatcaggt gctgctgcat ggaattcttg ggacaaaaag 1140

aacatagaaa ctgactcaga acctgctgct tggggttctc agggaaaaaa gaattctgaa 1200

actgagtcag gtcctgctgc ttggggtgct tgggacaaaa agaaatcgga aactgaacca 1260

ggtcctgctg gttggggtat gggggacaaa aagaattctg aaactgaatt aggccctgct 1320

gctatgggta attgggacaa aaagaaatcg gatacgaaat caggtccagc tgcttgggga 1380

tctacagatg ctgctgcttg gggttctagt gacaaaaaca actcggaaac tgaatcagat 1440

gctgctgctt ggggttctag gaacaaaaag acttctgaaa ttgaatcagg tgctggcgct 1500

tggggttcgt ggggacagcc aagtcccact gcagaagaca aggataccaa tgaggatgat 1560

agaaatccgt gggtttcctt gaaggagacc aaaagtagag agaaggatga taaagagaga 1620

agccagtggg gaaatccagc caagaaattt ccatctagtg gtggctggtc caatggtggt 1680

ggtgcagact ggaaaggaaa taggaaccat actcctaggc cacctagatc tgaggataat 1740

ttggctccca tgtttacagc aacgagacag cgcttggact cattcacctc tgaggagcaa 1800

gagctgctct cagatgttga gccagtcatg aggactctca ggaaaattat gcatccgtct 1860

gcgtaccctg atggagaccc aatttctgat gatgacaaga cgtttgtcct ggagaagatt 1920

ttgaacttcc atccccaaaa agagacaaag cttggttctg gcgttgactt catcacggtt 1980

gataagcaca ccatcttctc agacagcagg tgcttctttg ttgtctcaac tgatggggcc 2040

aaacaggact tttcataccg aaaatccctc aacaactatt tgatgaagaa ataccctgac 2100

cgtgctgaag aattcattga caagtacttt acgaaacctc ggcctagcgg aaacagagac 2160

aggaacaacc aagatgccac accacctggg gaagaacagt ctcagcctcc gaatcagtct 2220

ataggtaatg gaggtgacga ctttcagaca cagactcagt ctcagtctcc atctcagact 2280

cgggctcagt ctccatctca ggctcaggct cagtctccat ctcagacgca gtctcagtct 2340

cagtctcagt ctcagtctca gtctcagtct cagtctcagt ctcagtctca gtctcagtct 2400

cagtctcagt ctcagtctca gtctcagtct ccatcccaga ctcagactca gtctccgtct 2460

cagactcagg ctcaggctca gtctccatca tctcagtctc cgtcccagac gcagacataa 2520

<210> 2
<211> 839
<212> PRT
<213> Arabidopsis thaliana

<400> 2

Met Val Ser Lys Gly Val Leu Lys Glu His Ile Ile Leu Leu Ala Asn
1 5 10 15

Asn Met Thr Cys Ser Gly Thr Met Leu Gly Phe Asn Ser Gly Gly Tyr
20 25 30

Lys Ala Leu Thr Arg Ser Leu Asn Ile Lys Ala Pro Phe Thr Glu Ala
35 40 45

Thr Leu Ile Ala Pro Arg Lys Cys Phe Glu Lys Ala Ala Glu Lys Cys
50 55 60

His Thr Asp Ser Leu Ser Thr Val Val Gly Ser Cys Ser Trp Gly Lys
65 70 75 80

Arg Val Asp Val Gly Thr Gly Ser Gln Phe Glu Leu Leu Trp Asn Gln
85 90 95

Lys Glu Thr Gly Leu Asp Asp Lys Glu Glu Thr Asp Val Tyr Ser Phe
100 105 110

Leu Gln Met Val Ile Ser Thr Thr Asn Ala Asp Ala Phe Val Ser Ser
115 120 125

Pro Gly Phe Asp Val Thr Glu Glu Glu Met Ala Glu Trp Ala Glu Ser
130 135 140

Pro Glu Arg Asp Ser Ala Leu Gly Glu Pro Lys Phe Glu Asp Ser Ala
145 150 155 160

Asp Phe Gln Asn Leu His Asp Glu Gly Lys Pro Ser Gly Ala Asn Trp
165 170 175

Glu Lys Ser Ser Ser Trp Asp Asn Gly Cys Ser Gly Gly Ser Glu Trp
180 185 190

Gly Val Ser Lys Ser Thr Gly Gly Glu Ala Asn Pro Glu Ser Asn Trp
195 200 205

Glu Lys Thr Thr Asn Val Glu Lys Glu Asp Ala Trp Ser Ser Trp Asn
210 215 220

Thr Arg Lys Asp Ala Gln Glu Ser Ser Lys Ser Asp Ser Gly Gly Ala
225 230 235 240

Trp Gly Ile Lys Thr Lys Asp Ala Asp Ala Asp Thr Thr Pro Asn Trp
245 250 255

Glu Thr Ser Pro Ala Pro Lys Asp Ser Ile Val Pro Glu Asn Asn Glu
260 265 270

Pro Thr Ser Asp Val Trp Gly His Lys Ser Val Ser Asp Lys Ser Trp
275 280 285

Asp Lys Lys Asn Trp Gly Thr Glu Ser Ala Pro Ala Ala Trp Gly Ser

290 295 300

Thr Asp Ala Ala Val Trp Gly Ser Ser Asp Lys Lys Asn Ser Glu Thr
305 310 315 320

Glu Ser Asp Ala Ala Ala Trp Gly Ser Arg Asp Lys Asn Asn Ser Asp
325 330 335

Val Gly Ser Gly Ala Gly Val Leu Gly Pro Trp Asn Lys Lys Ser Ser
340 345 350

Glu Thr Glu Ser Asn Gly Ala Thr Trp Gly Ser Ser Asp Lys Thr Lys
355 360 365

Ser Gly Ala Ala Ala Trp Asn Ser Trp Asp Lys Lys Asn Ile Glu Thr
370 375 380

Asp Ser Glu Pro Ala Ala Trp Gly Ser Gln Gly Lys Lys Asn Ser Glu
385 390 395 400

Thr Glu Ser Gly Pro Ala Ala Trp Gly Ala Trp Asp Lys Lys Lys Ser
405 410 415

Glu Thr Glu Pro Gly Pro Ala Gly Trp Gly Met Gly Asp Lys Lys Asn
420 425 430

Ser Glu Thr Glu Leu Gly Pro Ala Ala Met Gly Asn Trp Asp Lys Lys
435 440 445

Lys Ser Asp Thr Lys Ser Gly Pro Ala Ala Trp Gly Ser Thr Asp Ala
450 455 460

Ala Ala Trp Gly Ser Ser Asp Lys Asn Asn Ser Glu Thr Glu Ser Asp
465 470 475 480

Ala Ala Ala Trp Gly Ser Arg Asn Lys Lys Thr Ser Glu Ile Glu Ser
485 490 495

Gly Ala Gly Ala Trp Gly Ser Trp Gly Gln Pro Ser Pro Thr Ala Glu
500 505 510

Asp Lys Asp Thr Asn Glu Asp Asp Arg Asn Pro Trp Val Ser Leu Lys
515 520 525

Glu Thr Lys Ser Arg Glu Lys Asp Asp Lys Glu Arg Ser Gln Trp Gly
530 535 540

```
Asn Pro Ala Lys Lys Phe Pro Ser Ser Gly Gly Trp Ser Asn Gly Gly
545                 550                 555                 560

Gly Ala Asp Trp Lys Gly Asn Arg Asn His Thr Pro Arg Pro Pro Arg
                565                 570                 575

Ser Glu Asp Asn Leu Ala Pro Met Phe Thr Ala Thr Arg Gln Arg Leu
            580                 585                 590

Asp Ser Phe Thr Ser Glu Glu Gln Glu Leu Leu Ser Asp Val Glu Pro
        595                 600                 605

Val Met Arg Thr Leu Arg Lys Ile Met His Pro Ser Ala Tyr Pro Asp
    610                 615                 620

Gly Asp Pro Ile Ser Asp Asp Asp Lys Thr Phe Val Leu Glu Lys Ile
625                 630                 635                 640

Leu Asn Phe His Pro Gln Lys Glu Thr Lys Leu Gly Ser Gly Val Asp
                645                 650                 655

Phe Ile Thr Val Asp Lys His Thr Ile Phe Ser Asp Ser Arg Cys Phe
            660                 665                 670

Phe Val Val Ser Thr Asp Gly Ala Lys Gln Asp Phe Ser Tyr Arg Lys
        675                 680                 685

Ser Leu Asn Asn Tyr Leu Met Lys Lys Tyr Pro Asp Arg Ala Glu Glu
    690                 695                 700

Phe Ile Asp Lys Tyr Phe Thr Lys Pro Arg Pro Ser Gly Asn Arg Asp
705                 710                 715                 720

Arg Asn Asn Gln Asp Ala Thr Pro Pro Gly Glu Glu Gln Ser Gln Pro
                725                 730                 735

Pro Asn Gln Ser Ile Gly Asn Gly Gly Asp Asp Phe Gln Thr Gln Thr
            740                 745                 750

Gln Ser Gln Ser Pro Ser Gln Thr Arg Ala Gln Ser Pro Ser Gln Ala
        755                 760                 765

Gln Ala Gln Ser Pro Ser Gln Thr Gln Ser Gln Ser Gln Ser Gln Ser
    770                 775                 780

Gln Ser Gln Ser Gln Ser Gln Ser Gln Ser Gln Ser Gln Ser Gln Ser
```

785 790 795 800

Gln Ser Gln Ser Gln Ser Gln Ser Gln Ser Pro Ser Gln Thr Gln Thr
805 810 815

Gln Ser Pro Ser Gln Thr Gln Ala Gln Ala Gln Ser Pro Ser Ser Gln
820 825 830

Ser Pro Ser Gln Thr Gln Thr
835

<210> 3
<211> 535
<212> DNA
<213> Arabidopsis thaliana

<400> 3
ttttttttttt ttttttccaa actataaaaa tcaatgtatt caaattactt ttgggtttgc 60

ttcatgttcg cttgaaatgt tctcgtataa atctgtcggc atgagacggg tacttttccc 120

ggatgtatgc acggaggcac ttctggtatg agaaatcaat ccaaacaccg tctgttcgaa 180

caacgaagag gcatcttgac tttctgaact cagggtgcct atcaaccatt atagcgtcaa 240

gtccacaccc aattttgtcc tgcgcgcgtg gatggtagga gaggagcttc tccacgacga 300

cattttcgtc ccttgggcag agacattcgc catccccgta tctatcggag tggaggatat 360

ccttgatgag caggacgacg gggtcaatat cgcgaaggat ctccgcctcc acggccctaa 420

acggcagcgg gggctccggc gtatccacaa ccggcggggc ggcgagcttc caatttggcg 480

ggagcgccgg gggcggcgga tcaccggcag tgggaggccc agagagaagt cgacg 535

<210> 4
<211> 694
<212> DNA
<213> Arabidopsis thaliana

<400> 4
gaaagagtga gcactgataa cttgatcgtt acaatgctcc aatttctgct ttcctaacag 60

caaattgcag cactagcaag taactagcag ccaaaaaaaa aaatacaggg catggccatg 120

gccatggctt cgttatctaa accaccacca ccgctatccc ttcacgcttt gtcaaaccct 180

atctctctat tttcctcgcc ggcggtggtg ttatcttttc cgtttaatcg aacagctttc 240

gcgagtccaa gtgcgcgcgc aattaagaca ggatcggaca gttcagattt gttgaggaaa 300

ccgattgcac cgtcagagaa ggacttgttt ggaatttctg aagaagaaga ggaggaggac 360

agtgaagggg aaagggagga agaagaagag gaaggatttg tggattggga ggataggatt 420

ttggaggata ccgtcccact tgttgggttt gttagaatga ttcttcattc tgacaaatat 480

gaaagtggag acagattgag tcaagagcat gagagaacta ttgtggatag attgcttgct 540

catcatccag attttgataa caagattgga tgtggaattg attatattac agtcaggtat 600

catcctgatt ttgcggactc tcggtgtttg ttcatagttc gaaaagatgg gcaatcggtt 660

gacttttctt attggaaatg cctaaagggc ttga 694

<210> 5
<211> 1057
<212> DNA
<213> Arabidopsis thaliana

<400> 5
gctttagaca aaattaaagt ttataaccca aaagagtgaa aaaagaatcc aagagggcat 60

aagtcaaaaa ctactgatca ctcaaaaaac gcttaacaat ggctgctccg ttactcctca 120

gaggtctacc acttctccgg cttcgcttct gctatcagcg gctgccgaat cctagctttg 180

ttcaactttc tccgaggctg ttttcagctg cttccgaatc ggctcagccc ttcgacgacg 240

cggcgtccga tggggaccag accaccgccg tgctaagcgc caaagagccg cccaattacc 300

cgagatggag tgaccccgac tatcgaaaat ggaaagacca ggaaactcag atactcaaag 360

atatcgagcc cgtaattttc ctggctaagg aaattatcca ctccgatagg tatatggatg 420

gagaacgttt gacagcagaa gatgagaaaa tagtggtaga caagcttctt gcttatcatc 480

cccattctga agataaaatt ggatgtggcc tcgactcaat tatggttgat cggcatcccc 540

agttcaaacg atccaggtgt ctctttgtcg taagattgga tggtggatgg atagactttt 600

cctatcagaa gtgtcttaga cagtacattc gagataagta cccttcttat gccgaaaaat 660

tcataaaaga acacttcaaa cgtggtagtt gagttttatt tcaaacagcc tctcctcaca 720

taaagactaa agttggaagt tgttggcctg taagagccct tatggtgtgt ctttctcaat 780

gaaggaacaa gaaaatgaca ggatcttatg aatgcttggt gtgcgactaa aggtgtttat 840

ggattaaatg accaacaaat gactctctga tacttttagt caatgtggta gatattagtt 900

gaccagagca attcaagttt tatttggtgg gagaataaag ctgatggaaa tttaaaaaaa 960

aagtacactt ccatagtaat tagatggctc tgcttgtaat gttgactatc actcttttt 1020

ggtgaatcac tgtctctatc ttctttaatt tctgaaa 1057

<210> 6
<211> 935
<212> DNA
<213> Arabidopsis thaliana

<400> 6
ggctgctccg ttactcctca gaggtctgcc aatcctccgg ctccgcttct gctatcagcg 60

gctgccgaat cctagctttg ttctactttc tccgaggctg ttttcagctg cttccgaatc 120

ggctcagccc ttcgacgacg cggcgtccga tggggaccag accaccgccg tgctaagcgc 180

caaagacccg cccaattacc cgagatggac tgacccggac tatcgaaaat ggaaagacca 240

ggaaactcag atactcaaag atatcgagcc cgttattttc ctggccaagg aaattatcca 300

ctccgatagg tatatggatg gagaacgtct gacagcagaa gatgagaaaa tagtggtaga 360

caagcttctt gcttatcatc cccattctga agataaaatt ggatgtggcc tcgactcaat 420

tatggttgat cggcatcccc agttcaaacg atccaggtgt ctctttgtcg taagattgga 480

cggtggatgg atagactttt cctatcagaa gtgtcttagg cagtacattc gagataagta 540

cccttcttat gctgaaaaat tcataaaaga acacttcaaa cgtggtagtt gagtttcatt 600

tcaaacaacc tctcctcaca taaatactaa agttggaagc tgttggcctg taagagccct 660

catggtgggt ctttctcaat gaaggaacaa gaaaatgaca ggaacttatg aatgcttggt 720

gtgtggctaa aggtgtttat ggattaaatg accaacaaat gactttctga tacctttagt 780

caatgtggta gatattagtt gaccagagca tttcaagttt tatttggtgg gagaataaag 840

ttgatggaaa taaaaaaaaa atgtacactt ccatagtaat tagttggctt tgcttgtaat 900

gttgactatc actctttttt ggtgaatcac tgtct 935