

(19) World Intellectual Property
Organization
International Bureau



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

PCT

(10) International Publication Number
WO 2004/093532 A2

- (51) International Patent Classification⁷: **A01H**
- (21) International Application Number: PCT/US2004/012279
- (22) International Filing Date: 19 April 2004 (19.04.2004)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/464,487 22 April 2003 (22.04.2003) US
- (71) Applicant (for all designated States except US): **AGRI-NOMICS LLC** [US/US]; 16160 S.W. Upper Boones Ferry Road, Portland, Oregon 97224-7744 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **LIGHTNER, Jonathan** [US/US]; 28750 S. Heisinger Lane, Mulino, Oregon 97042 (US). **NG, Hein, Tsoeng** [CN/US]; 12975 SW 19th Street, Beaverton, Oregon 97008 (US).
- (74) Agents: **SALE, Terri, L.** et al.; P.O. Box 511, 170 Harbor Way, South San Francisco, California 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.*



WO 2004/093532 A2

(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 HIO106.2 nucleic acid. The invention is further directed to methods of generating plants with an altered oil content phenotype.

GENERATION OF PLANTS WITH ALTERED OIL CONTENT

REFERENCE TO RELATED APPLICATIONS

This application claims priority to U.S. provisional patent application 60/464,487
5 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,
10 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,
15 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
20 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,
25 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
30 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
5 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
10 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
15 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*
20 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
25 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
30 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
5 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.

10 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).

15 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 HIO106.2 polypeptide. In preferred embodiments, the transgenic plant is selected from the group consisting of
20 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
25 nucleotide sequence that encodes or is complementary to a sequence that encodes a HIO106.2 polypeptide, and growing the transformed progenitor cells to produce a transgenic plant, wherein the HIO106.2 polynucleotide sequence is expressed causing the high oil phenotype.

30 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-
5 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
10 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).

15 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
20 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
25 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
30 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
5 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
10 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
15 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,
20 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
25 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
30 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 "HIO106.2," (At2g34220; GI#18403485) encoding a non-secretory protein that lacks a transmembrane domain (GI#15226250), 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 38.4% relative to a planting day average oil content of 35.5% (108% of PDA). The association of the HIO106.2 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, HIO106.2 genes and/or polypeptides may be employed in the development of genetically modified plants having a modified oil content phenotype ("a HIO106.2 phenotype"). HIO106.2 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. HIO106.2 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 HIO106.2 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.

30

HIO106.2 Nucleic Acids and Polypeptides

Arabidopsis HIO106.2 nucleic acid (genomic DNA) sequence is provided in SEQ ID NO:1 and in Genbank entry GI#18403485. The corresponding protein sequence is

provided in SEQ ID NO:2 and in GI#15226250. Nucleic acids and/or proteins that are orthologs or paralogs of *Arabidopsis* HIO106.2, are described in Example 3 below.

As used herein, the term "HIO106.2 polypeptide" refers to a full-length HIO106.2 protein or a fragment, derivative (variant), or ortholog thereof that is "functionally active,"
5 meaning that the protein fragment, derivative, or ortholog exhibits one or more of the functional activities associated with the polypeptide of SEQ ID NO:2. In one preferred embodiment, a functionally active HIO106.2 polypeptide causes an altered oil content phenotype when mis-expressed in a plant. In a further preferred embodiment, mis-expression of the HIO106.2 polypeptide causes a high oil phenotype in a plant. In another
10 embodiment, a functionally active HIO106.2 polypeptide is capable of rescuing defective (including deficient) endogenous HIO106.2 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 HIO106.2 polypeptide (i.e., a native polypeptide having the sequence of SEQ ID
15 NO:2 or a naturally occurring ortholog thereof) retains one of more of the biological properties associated with the full-length HIO106.2 polypeptide, such as signaling activity, binding activity, catalytic activity, or cellular or extra-cellular localizing activity. A HIO106.2 fragment preferably comprises a HIO106.2 domain, such as a C- or N-terminal or catalytic domain, among others, and preferably comprises at least 10, preferably at least
20 20, more preferably at least 25, and most preferably at least 50 contiguous amino acids of a HIO106.2 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). A preferred HIO106.2 fragment comprises of one or more plant specific domains, including PF04780 and PF04781. Functionally active variants of full-length HIO106.2 polypeptides
25 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 HIO106.2 polypeptide. In some cases, variants are generated that change the post-translational processing of a HIO106.2 polypeptide. For instance, variants may have altered protein transport or protein localization characteristics or altered protein half-life
30 compared to the native polypeptide.

As used herein, the term "HIO106.2 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 HIO106.2

nucleic acid of this invention may be DNA, derived from genomic DNA or cDNA, or RNA.

In one embodiment, a functionally active HIO106.2 nucleic acid encodes or is complementary to a nucleic acid that encodes a functionally active HIO106.2 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 HIO106.2 polypeptide. A HIO106.2 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 HIO106.2 polypeptide, or an intermediate form. A HIO106.2 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 HIO106.2 nucleic acid is capable of being used in the generation of loss-of-function HIO106.2 phenotypes, for instance, via antisense suppression, co-suppression, etc.

In one preferred embodiment, a HIO106.2 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 HIO106.2 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 HIO106.2 polypeptide of the invention comprises a polypeptide sequence with at least 50% or 60% identity to the HIO106.2 polypeptide sequence of SEQ ID NO:2, and may have at least 70%, 80%, 85%, 90% or 95% or more sequence identity to the HIO106.2 polypeptide sequence of SEQ ID NO:2. In another embodiment, a HIO106.2 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, such as one or more plant specific domains, including PF04780 and PF04781. In yet another embodiment, a HIO106.2 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 and comprises of one or more plant specific domains, including PF04780 and PF04781.

In another aspect, a HIO106.2 polynucleotide sequence is at least 50% to 60% identical over its entire length to the HIO106.2 nucleic acid sequence presented as SEQ ID NO:1, or nucleic acid sequences that are complementary to such a HIO106.2 sequence, and may comprise at least 70%, 80%, 85%, 90% or 95% or more sequence identity to the HIO106.2 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.*,
5 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
10 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
15 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
20 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
25 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 HIO106.2 polypeptide can be produced. For example, codons may be selected to increase the rate at which expression of the polypeptide occurs in a particular
30 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* HIO106.2. 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

5 "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

10 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

15 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

20 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

25 portion of the *Arabidopsis* HIO106.2 coding sequence may be used as a probe. HIO106.2 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

30 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 HIO106.2 polypeptides are used for ortholog isolation (see, e.g., Harlow and Lane, 1988, 1999). Western blot analysis can determine that a HIO106.2 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
5 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 HIO106.2 nucleic acid and/or polypeptide sequences have been identified.

HIO106.2 nucleic acids and polypeptides may be obtained using any available method. For instance, techniques for isolating cDNA or genomic DNA sequences of
10 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
15 the HIO106.2 nucleic acid into a plant expression vector for transformation of in plant cells, and the HIO106.2 polypeptide is expressed in the host plant.

An isolated HIO106.2 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
20 HIO106.2 nucleic acid. However, an isolated HIO106.2 nucleic acid molecule includes HIO106.2 nucleic acid molecules contained in cells that ordinarily express HIO106.2 where, for example, the nucleic acid molecule is in a chromosomal location different from that of natural cells.

25 **Generation of Genetically Modified Plants with an Altered Oil Content Phenotype**

HIO106.2 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,
30 altered expression of the HIO106.2 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 HIO106.2 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 annuus*), cotton (*Gossypium hirsutum*), corn (*Zea mays*), cocoa
5 (*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
10 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,
15 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
20 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 HIO106.2 polynucleotide may encode the entire protein or a biologically active portion thereof.

25 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
30 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).

5 Expression (including transcription and translation) of HIO106.2 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 HIO106.2 nucleic acid. These include constitutive, inducible and regulatable promoters,
10 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
15 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, HIO106.2 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
20 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 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
30 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 5 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).

10 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 HIO106.2 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 15 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 20 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).

25

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

30 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.

25 Generation of Mutated Plants with an Altered Oil Content Phenotype

The invention further provides a method of identifying plants that have mutations in endogenous HI0106.2 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 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. HI0106.2-specific PCR is used to identify whether a mutated plant has a HI0106.2 mutation. Plants having HI0106.2 mutations may then be tested for altered oil content, or alternatively, plants may be tested for altered oil content, and then HI0106.2-specific PCR is used to

determine whether a plant having altered oil content has a mutated HIO106.2 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).

5 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 HIO106.2 gene or orthologs of HIO106.2 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 HIO106.2 nucleic acid is
10 used to identify whether a plant having altered oil content has a mutation in endogenous HIO106.2 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
15 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.

20 EXAMPLES

EXAMPLE 1

Generation of Plants with a HIO106.2 Phenotype by Transformation with an Activation Tagging Construct

Mutants were generated using the activation tagging "ACTTAG" vector, pSKI015
25 (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
30 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).

5 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 $\geq 107\%$ 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

10 insertions, also displaying high oil, were identified and the flanking sequence confirmed by PCR and sequencing.

Line W000151578 (IN047787) had a NIR determined oil content of 37.7% relative to a planting day average oil content of 34.4% (110% of PDA). Line W000112024

15 (IN034777) had a NIR determined oil content of 38.4% relative to a planting day average oil content of 35.5% (108% of PDA).

Line W000151578 (IN047787) had a confirmed ACTTAG insertion on Chromosome 2 at bp 14400020.

20

Line W000112024 (IN034777) had a confirmed ACTTAG insertion on Chromosome 2 at bp 14407000.

EXAMPLE 2

25 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

30 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 IN047787 and IN034777, 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 IN047787, there was sequence identity to
5 nucleotides 21803-22156 on Arabidopsis clone F13P17 map ve016 chromosome 2 (GI#20197153), placing the left border junction downstream from 21803 (GI#20197153). The opposite flank (predicted right border junction) of this insert is not determined. Left border of IN047787 T-DNA is ~ 19 bp 5' of the translation start site.

For ACTTAG line IN034777, there was sequence identity to nucleotides 28624-
10 28769 on Arabidopsis clone F13P17 map ve016 chromosome 2 (GI#20197153), placing the left border junction downstream from 28624 (GI#20197153). The opposite flank (predicted right border junction) of this insert was not determined. Left border IN034777 T-DNA is ~ 6999 bp 3' of the translation start site.

15 EXAMPLE 3

Analysis of Arabidopsis HIO106.2 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
20 (Thompson JD *et al.*, 1994, Nucleic Acids Res 22:4673-4680).

TBLASTN against est_others results:

At2g34220 has limited homology to sequences in the EST databases (including Arabidopsis ESTs). It is likely that At2g34220 and its homologues are expressed at low
25 levels.

BLASTP against all.aa results:

The protein At2g34220 has homology to putative or hypothetical proteins from Arabidopsis. The top 10 BLAST results for At2g34220 are listed below and are included
30 in the "Orthologue Table."

1. At2g34220 itself (several redundant entries)

>gi|15226250|ref|NP_180969.1| hypothetical protein; protein id: At2g34220.1
 [Arabidopsis thaliana] >gi|7485475|pir|T02310 hypothetical protein At2g34220
 [imported] - Arabidopsis thaliana >gi|3337353|gb|AAC27398.1| hypothetical protein
 5 [Arabidopsis thaliana]
 Score = 3506 (1239.2 bits), Expect = 0., P = 0.

2. At2g34240 from Arabidopsis (several redundant entries)

>gi|15226253|ref|NP_180971.1| hypothetical protein; protein id: At2g34240.1
 10 [Arabidopsis thaliana] >gi|7485477|pir|T02312 hypothetical protein At2g34240
 [imported] - Arabidopsis thaliana >gi|3337355|gb|AAC27400.1| hypothetical protein
 [Arabidopsis thaliana]
 Score = 2286 (809.8 bits), Expect = 1.6e-236, P = 1.6e-236

3. At2g34230 from Arabidopsis (several redundant entries)

>gi|15226252|ref|NP_180970.1| hypothetical protein; protein id: At2g34230.1
 [Arabidopsis thaliana] >gi|7485476|pir|T02311 hypothetical protein At2g34230
 [imported] - Arabidopsis thaliana >gi|3337354|gb|AAC27399.1| hypothetical protein
 [Arabidopsis thaliana]
 20 Score = 2219 (786.2 bits), Expect = 2.0e-229, P = 2.0e-229

4. At1g65130 from Arabidopsis (several redundant entries)

>gi|15217861|ref|NP_176692.1| hypothetical protein; protein id: At1g65130.1
 [Arabidopsis thaliana]
 25 Score = 714 (256.4 bits), Expect = 2.1e-69, P = 2.1e-69

The following sequences are other redundant entries of At1g65130. 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
 30 activity.

>gi|25404528|pir|B96675 hypothetical protein T23K8.4 [imported] – Arabidopsis thaliana >gi|4646196|gb|AAD26869.1|AC007230_3 Contains similarity to F13P17.8
 gi|3337355 hypothetical protein from Arabidopsis thaliana BAC gb|AC004481

5. At1g65110 from Arabidopsis (several redundant entries)

>gi|15217858|ref|NP_176690.1| hypothetical protein; protein id: At1g65110.1
[Arabidopsis thaliana]

Score = 682 (245.1 bits), Expect = 7.4e-66, P = 7.4e-66

5

The following sequences are other redundant entries of At1g65110. 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.

10 >gi|25404524|pir||H96674 hypothetical protein T23K8.2 [imported] – Arabidopsis thaliana >gi|4646195|gb|AAD26868.1|AC007230_2 Contains similarity to F13P17.8 gi|3337355 hypothetical protein from Arabidopsis thaliana BAC gb|AC004481. EST gb|AA394560 comes from this gene

6. At1g65120 from Arabidopsis (several redundant entries)

>gi|15217860|ref|NP_176691.1| hypothetical protein; protein id: At1g65120.1
[Arabidopsis thaliana]

Score = 651 (234.2 bits), Expect = 2.6e-62, P = 2.6e-62

20 The following sequences are other redundant entries of At1g65120. 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|25404526|pir||A96675 hypothetical protein T23K8.3 [imported] – Arabidopsis thaliana >gi|4646208|gb|AAD26881.1|AC007230_15 Contains similarity to F13P17.8 gi|3337355 hypothetical protein from Arabidopsis thaliana BAC gb|AC004481

7. At1g52450 from Arabidopsis (several redundant entries)

>gi|15218999|ref|NP_175654.1| hypothetical protein; protein id: At1g52450.1
[Arabidopsis thaliana] >gi|25405552|pir||H96564 F6D8.33 [imported] - Arabidopsis

30 thaliana >gi|5903057|gb|AAD55616.1|AC008016_26 F6D8.33 [Arabidopsis thaliana]

Score = 643 (231.4 bits), Expect = 1.6e-61, P = 1.6e-61

8. At1g65200 from Arabidopsis

>gi|15217870|ref|NP_176699.1| hypothetical protein; protein id: At1g65200.1
 [Arabidopsis thaliana]

Score = 623 (224.4 bits), Expect = 2.0e-59, P = 2.0e-59

5

9. At1g52430 from Arabidopsis (several redundant entries)

>gi|15218996|ref|NP_175652.1| hypothetical protein; protein id: At1g52430.1

[Arabidopsis thaliana] >gi|25405550|pir||F96564 hypothetical protein F6D8.35 [imported]

- Arabidopsis thaliana >gi|5903058|gb|AAD55617.1|AC008016_27 Strong similarity to

10 F6D8.33. [Arabidopsis thaliana]

Score = 623 (224.4 bits), Expect = 2.3e-59, P = 2.3e-59

10. At1g52430 from Arabidopsis (2 redundant entries)

>gi|15240389|ref|NP_201000.1| putative protein; protein id: At5g61940.1

15 [Arabidopsis thaliana]

Score = 621 (223.7 bits), Expect = 3.2e-59, P = 3.2e-59

The following sequence is another redundant entry of At5g61940. However, it differs 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.

20

>gi|9758394|dbj|BAB08881.1| gb|AAD55616.1~gene_id:K22G18.7~strong similarity to unknown protein [Arabidopsis thaliana]

25

Closest Arabidopsis homologs:

At2g34240	Arabidopsis thaliana	>gi 15226253 >gi 7485477 >gi 3337355	Length = 712 Identities = 464/697 (66%), Positives = 538/697 (77%)	BLASTP Score = 2286 (809.8 bits), Expect = 1.6e-236, P = 1.6e-236
At2g34230	Arabidopsis thaliana	>gi 15226252 >gi 7485476 >gi 3337354	Length = 716 Identities = 449/703 (63%), Positives = 525/703 (74%)	BLASTP Score = 2219 (786.2 bits), Expect = 2.0e-229, P = 2.0e-229

At1g65130	Arabidopsis thaliana	>gi 15217861 >gi 25404528 >gi 4646196	Length = 1088 Identities = 205/694 (29%), Positives = 341/694 (49%)	BLASTP Score = 714 (256.4 bits), Expect = 2.1e-69, P = 2.1e-69
At1g65110	Arabidopsis thaliana	>gi 15217858 >gi 25404524 >gi 4646195	Length = 1094 Identities = 206/688 (29%), Positives = 332/688 (48%)	BLASTP Score = 682 (245.1 bits), Expect = 7.4e-66, P = 7.4e-66
At1g65120	Arabidopsis thaliana	>gi 15217860 >gi 25404526 >gi 4646208	Length = 1191 Identities = 206/714 (28%), Positives = 334/714 (46%)	BLASTP Score = 651 (234.2 bits), Expect = 2.6e-62, P = 2.6e-62
At1g52450	Arabidopsis thaliana	>gi 15218999 >gi 25405552 >gi 5903057	Length = 1136 Identities = 199/701 (28%), Positives = 331/701 (47%)	BLASTP Score = 643 (231.4 bits), Expect = 1.6e-61, P = 1.6e-61
At1g65200	Arabidopsis thaliana	>gi 15217870	Length = 1101 Identities = 192/684 (28%), Positives = 324/684 (47%)	BLASTP Score = 623 (224.4 bits), Expect = 2.0e-59, P = 2.0e-59
At1g52430	Arabidopsis thaliana	>gi 15218996 >gi 25405550 >gi 5903058	Length = 1136 Identities = 195/687 (28%), Positives = 318/687 (46%)	BLASTP Score = 623 (224.4 bits), Expect = 2.3e-59, P = 2.3e-59
At5g61940	Arabidopsis thaliana	>gi 15240389 >gi 9758394	Length = 1094 Identities = 200/708 (28%), Positives = 332/708 (46%)	BLASTP Score = 621 (223.7 bits), Expect = 3.2e-59, P = 3.2e-59

At2g34220 is a non-secretory protein that lacks a transmembrane domain (predicted by TMHMM) and signal peptide (predicted by SignalP). At2g34220 is likely to be localized to the nucleus or to the cytoplasm (44% cytoplasmic, 40% nuclear, 8% mitochondrial by Psort2).

Pfam analysis showed that At2g34220 contains two plant specific domains (PF04780, PF04781).

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
PF04781	1/1	18	144	1	126	246.1	5.6e-71
PF04780	1/1	192	647	1	468	866.0	1.4e-257

15

EXAMPLE 4

Generation of Mutated Plants with a HIO106.2 Phenotype

The invention further provides a method of identifying plants that have mutations in, or an allele of, endogenous HIO106.2 that confer a HIO106.2 phenotype, and
5 generating progeny of these plants that also have the HIO106.2 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. HIO106.2-specific PCR is used to identify whether a
10 mutated plant has a HIO106.2 mutation. Plants having HIO106.2 mutations may then be tested for the HIO106.2 phenotype, or alternatively, plants may be tested for the HIO106.2 phenotype, and then HIO106.2-specific PCR is used to determine whether a plant having the HIO106.2 phenotype has a mutated HIO106.2 gene. TILLING can identify mutations that may alter the expression of specific genes or the activity of proteins encoded by these
15 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 HIO106.2 gene or orthologs of HIO106.2 that may confer the HIO106.2 phenotype (see
20 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 HIO106.2 nucleic acid is used to identify whether a plant having a HIO106.2 phenotype has a mutation in endogenous
25 HIO106.2 or has a particular allele that causes the HIO106.2 phenotype compared to plants lacking the mutation or allele, and generating progeny of the identified plant that have inherited the HIO106.2 mutation or allele and have the HIO106.2 phenotype.

REFERENCES

- 30 Adams, M.D. *et al.*, *Science* 287(5461):2195, 2000.
Ageorges, A, *et al.*, *Plant Physiol. Biochem.* 38:177-185, 2000.
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.
- Ball, C. L. *et al.*, J. Cell. Sci. 108(Pt 8):2865-2875, 1995.
- 5 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.
- Chiang, P.W. *et al.*, Genomics 38(3): 421-424, 1996.
- 10 Chiang, P.W. *et al.*, Genomics 47(3):426-428, 1998.
- 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.
- 15 Davies, C., *et al.*, Plant Science 147:93-100, 1999.
- De Block *et al.*, Plant Physiol. (1989) 91:694-701.
- Dieffenbach C and Dveksler G (Eds.) PCR Primer: A Laboratory Manual, Cold Spring Harbor Laboratory Press, NY, 1989.
- Dornan, S. *et al.*, Mol. Biol. Cell 8 (8):1391-1403, 1997.
- 20 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.
- 25 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.
- 30 Glick, BR and Thompson, JE, Eds. Methods In Plant Molecular Biology and Biotechnology, p. 213-221, CRC Press, 1993.
- Gloeckner, G. *et al.*, Nature 418:79-85, 2002.
- Gonzalez-Gaitan, M. and Jackle, H. Cell 88(6): 767-776.
- Guo, S. *et al.*, Nature 408(6810):366-369, 2000.

- 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
- 5 Hayashi H *et al.*, *Science* 258: 1350-1353, 1992.
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.
- 10 Kaneko, T. *et al.*, *DNA Res.* 7:217-221, 2000.
Kaplan, C.D. *et al.*, *Genes Dev.* 14 (20):2623-2634, 2000.
Kardailsky I *et al.*, *Science* 286: 1962-1965, 1999.
Katavic V *et al.*, *Plant Physiology* 108(1):399-409, 1995.
Kawai, J. *et al.*, *Nature* 409(6821):685-690, 2001.
- 15 Keegan, B. R. *et al.*, *Nature* 408(6810):366-369, 2000.
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.
- 20 Lemieux B., *et al.*, 1990, *Theor Appl Genet* 80, 234-240.
Lavoie, S.B. *et al.*, *J Mol. Biol.* 312(4):675-685, 2001.
Ludwig, A., *et al.*, *Plant J.* 24:503-509, 2000.
Maeda, I., *et al.*, *Curr. Biol.* 11:171-176, 2001.
Meyer, S., *et al.*, *Plant J.* 24:869, 2000.
- 25 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.
Noiraud, N., *et al.*, *Plant Physiol.* 122:1447-1455, 2000.
Okuley *et al.*, *Plant Cell* 6(1):147-158, 1994.
- 30 Omirulleh *et al.*, *Plant Mol Biol.* 21(3): 415-28, 1993.
Robinson, M.S., *J. Cell. Biol.* 108(3):833-842, 1989.
Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual (Second Edition)*, Cold Sasaki, T., *Nature* 420:312-316, 2002.
Sauer, N. and Stolz, J., *Plant J.* 6:67-77, 1994.

- Scorilas, A. *et al.*, *Gene* 289(1-2):191-199, 2002.
- 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.
- 5 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.
- The C. elegans Sequencing Consortium *Science* 282(5396):2012-2018, 1998.
- Theologis, A. *et al.*, *Nature* 408:816-820, 2000.
- 10 Thompson JD *et al.*, *Nucleic Acids Res* 22:4673-4680, 1994.
- Tucker, K.L. *et al.*, *Nucleic Acids Res* 18(17):5306, 1990.
- 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.
- 15 Verdaguer B *et al.*, *Plant Mol Biol* 37:1055-1067, 1998.
- Voglmaier, S. M. *et al.*, *Biochem. Biophys. Res. Commun.* 187(1)L158-163, 1992.
- Wada, T. *et al.*, *Genes Dev* 12:343-356, 1998.
- Walker, J.C. *Plant J.* 3(3):45-456, 1993.
- Waterhouse, *et al.*, *Proc. Natl. Acad. Sci. USA* 95:13959-13964, 1998.
- 20 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.
- 25 Wood, V. *et al.*, *Nature* 415(6874):871-880, 2002.
- Yadav NS *et al.*, (1993) *Plant Physiol* 103, 467-476.
- Yamaguchi Y *et al.*, *Cell* 97:41-51, 1999.
- Yamaguchi Y *et al.*, *J Biol. Chem* 274:8085-8092, 1999.

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 HI0106.2 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 claim 1 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 HI0106.2 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 HI0106.2 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.

5

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

SEQUENCE LISTING

<110> Agrinomics LLC

<120> GENERATION OF PLANTS WITH ALTERED OIL CONTENT

<130> AG04-008C-PC

<150> 60/464,487

<151> 2003-04-22

<160> 2

<170> PatentIn version 3.2

<210> 1

<211> 2157

<212> DNA

<213> Arabidopsis thaliana

<400> 1

```

atgatgatga agagtacaga cgaaagtga acaaaatgga agatcgacga tgcagccgaa      60
cttgctagag acttctggct cgaacgtgac ggcatttcgg ctttgaaatt aacagagaaa      120
acgatttctg atcatgggaa tcacgaaaag tgctctagtc accacgaact gcaaggatgat      180
atcttctatg cccaagcgaa gaaaacagat gacaccgacc tcaaagtctt atacttgttt      240
gcttctgtcg atgcgatttc aatggctact cttctgtctc ctaacgcttt gactttgagg      300
tcctttcgtg gctatgctcg ttcgctaadc aagttgggtg atcagcttcg gataaacaag      360
ttttatgaga aagctgtgtc caaagcaaag ctgggaatgt ccataacgtc acctcaagga      420
aagtctacgg atgggtggcta ttttaccaag ataaagacag aacttcagaa ttggatagat      480
cttgcaacaa cgaagatgca ggcgactgtg gctgtttctg atataggtac tatggataat      540
cagaaagaac ctggggaaga agataacttct tttgatagat tgaagaagat ttgggttaaa      600
ttggatgaca agacccaaat aaactttttg gtagtagact ttgaaaagct tataggctat      660
atacaagata agcatgggac ggaagtaaaa gagcactttc agaaatgtat acccatcgcc      720
aataatcttc gatggagatg ttggaagtgc cacatttggt ctcagggtta ctactgtttt      780
acggattgca agatgcacat cttagataat catgtgcaca actctgaacc tgactttagt      840
gctcgtccta agtgtgtgga tgagattttg gctgacatga tatgctgtgg ggattggaag      900
cctgtggata cagcggaggc agaaaatctt atcaaggata gagctaagcg cggggaagag      960
tttgtttatg ttaatgggtg gtgtagtgat tggcctgaag cgaaagatcg cgagcgggaa     1020
aacgtactca aacagtttgc cgaggttcta aaatcttctt gtgccaaagga gaattgtact     1080
ctttcatgta ccctttggga ctggttgaga gattacacag aagagaatct ggaactacct     1140
ggagttcctg ggatctatct tgatcaatgt tgcttcttta agaaccctca atgcatctgc     1200
ttcttagatc ttaagcacct tgaacacatc ctcaagtatt ttagacagct caccacagat     1260

```

gttcgcgoga gtttagttcc aaaagtagtg aatcagtttt gggaaaattc ccagggtcaag 1320
 gaaagaatcg accttgaggg agtagctact tataatctgc ttctggacaa gagactgctg 1380
 tatgaagaag agcttgaatc ggataagata gggacagtag agcattataa atctacaggg 1440
 atttttgagg atgtaatgcc taaggggtgat aaaatcgtct cttggattct tgactgccca 1500
 gaaatcgatc ggggaatttgt gtctcaaatg gcaaacggcg tacacaatag taagctatgg 1560
 ttagcagctt tgagaatagt acggtgtgtg attaggaaaa aggaaagcta ctatgataag 1620
 agacggaaga tgctaaccta tgagaagatg ttgggtgagg ctgagacgat ttgtgacaga 1680
 gaagacacta ggaagaatgt taatcagcgg agcacatatg agtctgcact tcgaatgaaa 1740
 tgtgaggatc ttgtaggtaa gcaagatgat gatacaaaat gtttcttgag tgttgtaaga 1800
 gatgtttttg aaagacaaag ctctccgagc tttgaagtat tcgaagataa gaagtgtata 1860
 tccgaactct cttctactgt cccgaatgat gatgtgaaaa aaagtctggt aacgctaagg 1920
 aatcactga aagagaagtt ccctttgatt gattcaaaga ttttgcagaa caaatctacg 1980
 tacgaaaagc taatagatgt attcccgaag ctatcagcag ttgagtatcg cttgggtggtc 2040
 cttccattcg tgaagaagtt cttacaggat aaactgaaga aatgatgaa aacaaacagc 2100
 agttcagttg ctgcaggaga ttctgctgaa cggaaaccta agaagagaca tctttga 2157

<210> 2
 <211> 718
 <212> PRT
 <213> Arabidopsis thaliana

<400> 2

Met Met Met Lys Ser Thr Asp Glu Ser Glu Thr Lys Trp Lys Ile Asp
 1 5 10 15

Asp Ala Ala Glu Leu Ala Arg Asp Phe Trp Leu Glu Arg Asp Gly Ile
 20 25 30

Ser Ala Leu Lys Leu Thr Glu Lys Thr Ile Ser Asp His Gly Asn His
 35 40 45

Glu Lys Cys Ser Ser His His Glu Leu Gln Gly Asp Ile Phe Tyr Ala
 50 55 60

Gln Ala Lys Lys Thr Asp Asp Thr Asp Leu Lys Cys Leu Tyr Leu Phe
 65 70 75 80

Ala Ser Val Asp Ala Tyr Ser Met Ala Thr Leu Leu Ser Pro Asn Ala
 85 90 95

Leu Thr Leu Arg Ser Phe Arg Gly Tyr Ala Arg Ser Leu Ile Lys Leu
 100 105 110

Gly Asp Gln Leu Arg Ile Asn Lys Phe Tyr Glu Lys Ala Val Ser Lys
 115 120 125

Ala Lys Leu Gly Met Ser Ile Thr Ser Pro Gln Gly Lys Ser Thr Asp
 130 135 140

Gly Gly Tyr Phe Thr Lys Ile Lys Thr Glu Leu Gln Asn Trp Ile Asp
 145 150 155 160

Leu Ala Thr Thr Lys Met Gln Ala Thr Val Ala Val Ser Asp Ile Gly
 165 170 175

Thr Met Asp Asn Gln Lys Glu Pro Gly Glu Glu Asp Thr Ser Phe Asp
 180 185 190

Arg Leu Lys Lys Ile Trp Val Lys Leu Asp Asp Lys Thr Lys Ile Asn
 195 200 205

Phe Leu Val Val Asp Phe Glu Lys Leu Ile Gly Tyr Ile Gln Asp Lys
 210 215 220

His Gly Thr Glu Val Lys Glu His Phe Gln Lys Cys Ile Pro Ile Ala
 225 230 235 240

Asn Asn Leu Arg Trp Arg Cys Trp Lys Cys His Ile Cys Ser Gln Val
 245 250 255

Asn Tyr Cys Phe Thr Asp Cys Lys Met His Ile Leu Asp Asn His Val
 260 265 270

His Asn Ser Glu Pro Asp Phe Ser Ala Arg Pro Lys Cys Val Asp Glu
 275 280 285

Ile Leu Ala Asp Met Ile Cys Cys Gly Asp Trp Lys Pro Val Asp Thr
 290 295 300

Ala Glu Ala Glu Asn Leu Ile Lys Asp Arg Ala Lys Arg Gly Glu Glu
 305 310 315 320

Phe Val Tyr Val Asn Gly Trp Cys Ser Asp Trp Pro Glu Ala Lys Asp
 325 330 335

Arg Glu Arg Glu Asn Val Leu Lys Gln Phe Ala Glu Val Leu Lys Ser

			340						345							350
Ser	Cys	Ala	Lys	Glu	Asn	Cys	Thr	Leu	Ser	Cys	Thr	Leu	Trp	Asp	Trp	
		355						360					365			
Leu	Arg	Asp	Tyr	Thr	Glu	Glu	Asn	Leu	Glu	Leu	Pro	Gly	Val	Pro	Gly	
	370						375				380					
Ile	Tyr	Leu	Asp	Gln	Cys	Cys	Phe	Phe	Lys	Asn	Pro	Gln	Cys	Ile	Cys	
385					390					395					400	
Phe	Leu	Asp	Leu	Lys	His	Leu	Glu	His	Ile	Leu	Lys	Tyr	Phe	Arg	Gln	
				405					410					415		
Leu	Thr	Thr	Asp	Val	Arg	Ala	Ser	Leu	Val	Pro	Lys	Val	Val	Asn	Gln	
			420					425						430		
Phe	Trp	Glu	Asn	Ser	Gln	Val	Lys	Glu	Arg	Ile	Asp	Leu	Glu	Gly	Val	
		435					440					445				
Ala	Thr	Tyr	Asn	Leu	Leu	Leu	Asp	Lys	Arg	Leu	Leu	Tyr	Glu	Glu	Glu	
	450						455				460					
Leu	Glu	Ser	Asp	Lys	Ile	Gly	Thr	Val	Glu	His	Tyr	Lys	Ser	Thr	Gly	
465					470					475					480	
Ile	Phe	Glu	Asp	Val	Met	Pro	Lys	Gly	Asp	Lys	Ile	Val	Ser	Trp	Ile	
				485					490					495		
Leu	Asp	Cys	Pro	Glu	Ile	Asp	Arg	Glu	Phe	Val	Ser	Gln	Met	Ala	Asn	
			500					505					510			
Gly	Val	His	Asn	Ser	Lys	Leu	Trp	Leu	Ala	Ala	Leu	Arg	Ile	Val	Arg	
		515						520				525				
Cys	Val	Ile	Arg	Lys	Lys	Glu	Ser	Tyr	Tyr	Asp	Lys	Arg	Arg	Lys	Met	
	530					535					540					
Leu	Thr	Tyr	Glu	Lys	Met	Leu	Gly	Glu	Ala	Glu	Thr	Ile	Cys	Asp	Arg	
545					550					555					560	
Glu	Asp	Thr	Arg	Lys	Asn	Val	Asn	Gln	Arg	Ser	Thr	Tyr	Glu	Ser	Ala	
				565					570					575		
Leu	Arg	Met	Lys	Cys	Glu	Asp	Leu	Val	Gly	Lys	Gln	Asp	Asp	Asp	Thr	
		580						585						590		

Lys Cys Phe Leu Ser Val Val Arg Asp Val Phe Glu Arg Gln Ser Ser
595 600 605

Pro Ser Phe Glu Val Phe Glu Asp Lys Lys Cys Ile Ser Glu Leu Ser
610 615 620

Ser Thr Val Pro Asn Asp Asp Val Lys Lys Ser Leu Leu Thr Leu Arg
625 630 635 640

Lys Ser Leu Lys Glu Lys Phe Pro Leu Ile Asp Ser Lys Ile Leu Gln
645 650 655

Asn Lys Ser Thr Tyr Glu Lys Leu Ile Asp Val Phe Pro Lys Leu Ser
660 665 670

Ala Val Glu Tyr Arg Leu Val Val Leu Pro Phe Val Lys Lys Phe Leu
675 680 685

Gln Asp Lys Leu Lys Lys Met Met Lys Thr Asn Ser Ser Ser Val Ala
690 695 700

Ala Gly Asp Ser Ala Glu Arg Lys Pro Lys Lys Arg His Leu
705 710 715