

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
27 January 2005 (27.01.2005)

PCT

(10) International Publication Number
WO 2005/006846 A2

(51) International Patent Classification⁷: **A01H**
(21) International Application Number:
PCT/US2004/022578
(22) International Filing Date: 13 July 2004 (13.07.2004)
(25) Filing Language: English
(26) Publication Language: English

(30) Priority Data:
60/487,380 14 July 2003 (14.07.2003) 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.; Exelixis, Inc., 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, NA, 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 "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning 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 HIO110.3 nucleic acid. The invention is further directed to methods of generating plants with an altered oil content phenotype.

WO 2005/006846 A2

GENERATION OF PLANTS WITH ALTERED OIL CONTENT

REFERENCE TO RELATED APPLICATIONS

This application claims priority to U.S. provisional patent application 60/487,380
5 filed 7/14/2003, 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 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 about 3.5% to about 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 HIO110.3 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 HIO110.3 polypeptide, and growing the transformed progenitor cells to produce a transgenic plant, wherein the HIO110.3 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.

5 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
10 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
15 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
20 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
25 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
30 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 (ACTTAG) screen to identify the association between the gene we have designated "HIO110.3," (At1g47750; GI#30694149) encoding one or more non-secretory proteins (synonym: T2E6.18; GI#18401872), 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 tumefaciens*, 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 of pooled T3 progeny from the individual T2 plants. 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 39.8% relative to a planting day average oil content of 35.7% (111% of PDA). The association of the HIO110.3 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, HIO110.3 genes and/or polypeptides may be employed in the development of genetically modified plants having a modified oil content phenotype ("a HIO110.3 phenotype"). HIO110.3 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. HIO110.3 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 HIO110.3 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.

HIO110.3 Nucleic Acids and Polypeptides

10 *Arabidopsis* HIO110.3 nucleic acid (genomic DNA) sequence is provided in SEQ ID NO:1 and in Genbank entry GI#30694149. The corresponding protein sequence is provided in SEQ ID NO:2 and in GI#18401872. Nucleic acids and/or proteins that are orthologs or paralogs of *Arabidopsis* HIO110.3, are described in Example 3 below.

As used herein, the term "HIO110.3 polypeptide" refers to a full-length HIO110.3 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 of the functional activities associated with the polypeptide of SEQ ID NO:2. In one preferred embodiment, a functionally active HIO110.3 polypeptide causes an altered oil content phenotype when mis-expressed in a plant. In a further preferred embodiment, mis-expression of the HIO110.3 polypeptide causes a high oil phenotype in a plant. In another embodiment, a functionally active HIO110.3 polypeptide is capable of rescuing defective (including deficient) endogenous HIO110.3 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 HIO110.3 polypeptide (i.e., a native polypeptide having the sequence of SEQ ID NO:2 or a naturally occurring ortholog thereof) retains one or more of the biological properties associated with the full-length HIO110.3 polypeptide, such as signaling activity, binding activity, catalytic activity, or cellular or extra-cellular localizing activity. A HIO110.3 fragment preferably comprises a HIO110.3 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 HIO110.3 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 HIO110.3 fragment comprises of one or more non-secretory proteins (synonym: T2E6.18). Functionally active variants of full-length HIO110.3 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 HIO110.3 polypeptide. In some cases, variants are generated that change the post-translational processing of a HIO110.3 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 "HIO110.3 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 HIO110.3 nucleic acid of this invention may be DNA, derived from genomic DNA or cDNA, or RNA.

In one embodiment, a functionally active HIO110.3 nucleic acid encodes or is complementary to a nucleic acid that encodes a functionally active HIO110.3 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 HIO110.3 polypeptide. A HIO110.3 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 HIO110.3 polypeptide, or an intermediate form. A HIO110.3 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 HIO110.3 nucleic acid is capable of being used in the generation of loss-of-function HIO110.3 phenotypes, for instance, via antisense suppression, co-suppression, etc.

In one preferred embodiment, a HIO110.3 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 HIO110.3 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 HIO110.3 polypeptide of the invention comprises a polypeptide sequence with at least 50% or 60% identity to the HIO110.3 polypeptide sequence of SEQ ID NO:2, and may have at least 70%, 80%, 85%, 90% or 95% or more sequence identity to the HIO110.3 polypeptide sequence of SEQ ID NO:2, such as one or more non-secretory proteins (synonym: T2E6.18). In another embodiment, a HIO110.3 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 HIO110.3 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 non-secretory proteins (synonym: T2E6.18).

In another aspect, a HIO110.3 polynucleotide sequence is at least 50% to 60% identical over its entire length to the HIO110.3 nucleic acid sequence presented as SEQ ID NO:1, or nucleic acid sequences that are complementary to such a HIO110.3 sequence, and may comprise at least 70%, 80%, 85%, 90% or 95% or more sequence identity to the HIO110.3 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
5 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 HIO110.3 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
10 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* HIO110.3. 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
15 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
20 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,
25 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
30 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
5 portion of the *Arabidopsis* HIO110.3 coding sequence may be used as a probe. HIO110.3 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
10 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 HIO110.3 polypeptides are used for ortholog isolation (see, e.g., Harlow and Lane, 1988, 1999). Western blot analysis can determine that a HIO110.3 ortholog (i.e., an orthologous protein) is present in a crude extract of a particular plant species. When reactivity is
15 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
20 reverse BLAST against sequences from *Arabidopsis* or other species in which HIO110.3 nucleic acid and/or polypeptide sequences have been identified.

HIO110.3 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
25 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 HIO110.3 nucleic acid into a plant expression vector for transformation of in plant
30 cells, and the HIO110.3 polypeptide is expressed in the host plant.

An isolated HIO110.3 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

HIO110.3 nucleic acid. However, an isolated HIO110.3 nucleic acid molecule includes HIO110.3 nucleic acid molecules contained in cells that ordinarily express HIO110.3 where, for example, the nucleic acid molecule is in a chromosomal location different from that of natural cells.

5

Generation of Genetically Modified Plants with an Altered Oil Content Phenotype

HIO110.3 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; 10 the modified oil content is often observed in seeds. In a preferred embodiment, altered expression of the HIO110.3 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 HIO110.3 gene 15 (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 20 (*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* 25 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, 30 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
5 heterologous nucleic acid construct comprising an HIO110.3 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,
10 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
15 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
20 (Christou *et al.*, 1989; Kline *et al.*, 1987).

Expression (including transcription and translation) of HIO110.3 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
25 HIO110.3 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
30 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, HIO110.3 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 37:110-11; "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 HIO110.3 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.

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 HIO110.3 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 (about 4%) had high oil (defined as oil content of
5 $\geq 110\%$ 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 W000097868 (IN033255) had a NIR determined oil content of 39.8% relative
10 to a planting day average oil content of 35.7% (111% of PDA). Line W000142648 (IN044809) had a NIR determined oil content of 38.8% relative to a planting day average oil content of 36.1% (107% of PDA).

Line W000097868 (IN033255) had a confirmed ACTTAG insertion on Chromosome 1 at bp 17164913.

15 Line W000142648 (IN044809) had a confirmed ACTTAG insertion on Chromosome 1 at bp 17165460.

EXAMPLE 2

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

20 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
25 vector, confirmed the presence of the 35S enhancer in plants from lines IN033255 and IN044809, 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
30 search of the *Arabidopsis* Information Resource (TAIR) database (available at the arabidopsis.org website). For ACTTAG line IN033255, there was sequence identity to nucleotides 64283- 64817 on Arabidopsis BAC clone T2E6 chromosome 1 (GI#9743359), placing the left border junction upstream from nucleotide 64817 (GI#9743359). The

opposite flank (predicted right border junction) of this insert was not determined. Left border of IN033255 T-DNA was about 8448 bp 5' of the translation start site.

For ACTTAG line IN044809, there was sequence identity to nucleotides 63685-64270 on Arabidopsis BAC clone T2E6 chromosome 1 (GI#9743359), placing the left
5 border junction downstream from nucleotide 63685 (GI#9743359). The opposite flank (predicted right border junction) of this insert was not determined. Left border IN044809 T-DNA is about 9578 bp 5' of the translation start site.

EXAMPLE 3

10 Analysis of Arabidopsis HIO110.3 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).

15

TBLASTN against est_others results:

The candidate gene At1g47750 is supported by full length cDNA (gi:30694149).

There are many ESTs from diverse plant species showing similarity to At1g47750.
20 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*, *Mentha x piperita*, *Populus tremula*, *Oryza sativa*, *Lycopersicon esculentum*, *Solanum tuberosum*, and *Beta vulgaris*.

25

1. One EST contig from wheat

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

2. One EST contig from cotton

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

30

3. One EST contig from maize

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

4. One EST contig from soybean

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

5. One EST contig from poplars

5 The contigged sequence is presented as SEQ ID NO: 7.

6. One EST from rice

>gi|18388132 BM421330 V005F08 Oryza sativa mature leaf library induced by
M.grisea Oryzasativa cDNA clone V005F08

10

7. One EST from tomato

> gi|6058991 EST286576 tomato mixed elicitor, BTI Lycopersicon esculentum
cDNA clone cLET24B12, MRNA sequence

15

8. One EST contig from potato

The contigged sequence is presented as SEQ ID NO: 8.

9. One EST contig from sugar beet

The contigged sequence is presented as SEQ ID NO: 9.

20

BLASTP against all.aa results:

The protein At1g47750 has homology to a number of proteins from plants. The top
10 BLAST results for At1g47750 are listed below and are included in the Ortholog Table.

25

1. Itself (several redundant entries)

>gi|18401872|ref|NP_564514.1| expressed protein [Arabidopsis thaliana]
>gi|25405892|pir||B96518 protein T2E6.18 [imported] – Arabidopsis thaliana
>gi|9802590|gb|AAF99792.1|AC012463_9 T2E6.18 [Arabidopsis thaliana]
Score = 1166 (415.5 bits), Expect = 8.0e-118, P = 8.0e-118

30

2. At3g47430 from Arabidopsis (5 redundant entries)

>gi|15232797|ref|NP_190327.1| expressed protein [Arabidopsis thaliana]
>gi|7487306|pir||T12998 hypothetical protein T21L8.180 - Arabidopsis thaliana

>gi|5541710|emb|CAB51215.1| putative protein [Arabidopsis thaliana]
 >gi|25083111|gb|AAN72044.1| putative protein [Arabidopsis thaliana]
 >gi|30023692|gb|AAP13379.1| At3g47430 [Arabidopsis thaliana]
 Score = 229 (85.7 bits), Expect = 1.6e-18, P = 1.6e-18

5

3. At2g45740 from Arabidopsis (several redundant entries)

>gi|18406877|ref|NP_566055.1| expressed protein [Arabidopsis thaliana]
 >gi|30690116|ref|NP_850441.1| expressed protein [Arabidopsis thaliana]
 >gi|15450880|gb|AAK96711.1| Unknown protein [Arabidopsis thaliana]
 >gi|20197204|gb|AAC28551.2| expressed protein [Arabidopsis thaliana]
 >gi|21537163|gb|AAM61504.1| unknown [Arabidopsis thaliana]
 Score = 132 (51.5 bits), Expect = 2.7e-06, P = 2.7e-06

10

4. An unknown protein from tomato

>gi|8489788|gb|AAF75750.1|AF261140_1 unknown [Lycopersicon esculentum]
 Score = 131 (51.2 bits), Expect = 3.6e-06, P = 3.6e-06

15

5. Peroxisomal membrane protein PMP30B from yeast (3 redundant entries)

>gi|2498762|sp|Q00317|PXBB_CANBO PEROXISOMAL MEMBRANE
 PROTEIN PMP30B (PMP32) (PEROXIN-11B) >gi|1076947|pir|S50281 PMP32
 protein – yeast (Candida boidinii) >gi|457393|gb|AAA66347.1| peroxisomal
 membrane protein 32
 Score = 122 (48.0 bits), Expect = 6.2e-05, P = 6.2e-05

20

6. At3g61070 from Arabidopsis (several redundant entries)

>gi|30695340|ref|NP_191666.2| expressed protein [Arabidopsis thaliana]
 >gi|27754657|gb|AAO22773.1| unknown protein [Arabidopsis thaliana]
 >gi|28394051|gb|AAO42433.1| unknown protein [Arabidopsis thaliana]
 Score = 120 (47.3 bits), Expect = 7.9e-05, P = 7.9e-05

25

30

7. A predicted protein from rice

>gi|6983876|dbj|BAA90811.1| ESTs AU064537(E31904),AU082147(E31904)
 correspond to a region of the predicted gene. Similar to Arabidopsis thaliana

chromosome II BAC F4I18 genomic sequence; unknown protein. (AC004665)
 [Oryza sativa (japonica cultivar-group)]
 Score = 119 (46.9 bits), Expect = 0.00014, P = 0.00014

5 **8. At1g01820 from Arabidopsis (several redundant entries)**

>gi|18378887|ref|NP_563636.1| expressed protein [Arabidopsis thaliana]
 >gi|25511713|pir||A86150 T1N6.24 protein - Arabidopsis thaliana
 >gi|8671852|gb|AAF78415.1|AC009273_21 Contains similarity to an unknown
 protein F4I18.28 gi|7486466 from Arabidopsis thaliana BAC F4I18 gb|AC004665.
 10 ESTs gb|F14309, gb|AI998750, gb|995247, gb|T14224 and gb|AI995247 come
 from this gene >gi|12083290|gb|AAG48804.1|AF332441_1 unknown protein
 [Arabidopsis thaliana] >gi|17381255|gb|AAL36046.1| At1g01820/T1N6_18
 [Arabidopsis thaliana] >gi|20453367|gb|AAM19922.1| At1g01820/T1N6_18
 [Arabidopsis thaliana] >gi|21555588|gb|AAM63892.1| unknown [Arabidopsis
 15 thaliana]
 Score = 118 (46.6 bits), Expect = 0.00015, P = 0.00015

9. A predicted protein from rice

>gi|25553559|dbj|BAC24830.1| OSJNBa0038F22.7 [Oryza sativa (japonica
 20 cultivar-group)]
 Score = 119 (46.9 bits), Expect = 0.00039, P = 0.00039

10. Peroxisomal membrane protein PMP30A from yeast (3 redundant entries)

>gi|2498761|sp|Q00316|PXBA_CANBO PEROXISOMAL MEMBRANE
 25 PROTEIN PMP30A (PMP31) (PEROXIN-11A) >gi|1076946|pir||S50280 PMP31
 protein – yeast (Candida boidinii) >gi|457391|gb|AAA66346.1| peroxisomal
 membrane protein 31
 Score = 113 (44.8 bits), Expect = 0.00074, P = 0.00074

30

Ortholog Gene Name	Species	GI #	% ID to HIO110.3	Score(s) (BLAST, Clustal, etc.)
An unknown protein from tomato	<i>Lycopersicon esculentum</i>	> gi 8489788 	Length = 235 Identities = 39/142 (27%), Positives = 75/142 (52%)	BLASTP Score = 131 (51.2 bits), Expect = 3.6e-06, P = 3.6e-06
Peroxisomal membrane protein PMP30B from yeast	<i>Candida boidinii</i>	> gi 2498762 > gi 1076947 > gi 457393 	Length = 256 Identities = 51/191 (26%), Positives = 90/191 (47%)	BLASTP Score = 122 (48.0 bits), Expect = 6.2e-05, P = 6.2e-05
A predicted protein from rice	<i>Oryza sativa (japonica cultivar-group)</i>	> gi 6983876 	Length = 252 Identities = 36/142 (25%), Positives = 72/142 (50%)	BLASTP Score = 119 (46.9 bits), Expect = 0.00014, P = 0.00014
A predicted protein from rice	<i>Oryza sativa (japonica cultivar-group)</i>	> gi 25553559 	Length = 422 Identities = 36/142 (25%), Positives = 72/142 (50%)	BLASTP Score = 119 (46.9 bits), Expect = 0.00039, P = 0.00039
Peroxisomal membrane protein PMP30B from yeast	<i>Candida boidinii</i>	> gi 2498761 > gi 1076946 > gi 457391 	Length = 256 Identities = 52/191 (27%), Positives = 90/191 (47%)	BLASTP Score = 113 (44.8 bits), Expect = 0.00074, P = 0.00074
One EST contig from wheat	<i>Triticum aestivum</i>	gi 17145812 gi 21997469 	Length = 602 Identities = 67/154 (43%), Positives = 97/154 (62%), Frame = +2	TBLASTN Score = 284 (105.0 bits), Expect = 9.0e-25, P = 9.0e-25
One EST contig from cotton	<i>Gossypium hirsutum</i>	gi 5049379 gi 5045140 	Length = 696 Identities = 34/139 (24%), Positives = 71/139 (51%), Frame = +2	TBLASTN Score = 109 (43.4 bits), Expect = 4.2e-05, P = 4.2e-05
One EST contig from maize	<i>Zea mays</i>	gi 19058211 gi 18964981 gi 19058273 gi 20300765 	Length = 572 Identities = 62/150 (41%), Positives = 92/150 (61%), Frame = +2	TBLASTN Score = 303 (111.7 bits), Expect = 4.0e-27, P = 4.0e-27

One EST contig from soybean	<i>Glycine max</i>	gil9440047 gil10237921 gil21602823	Length = 833 Identities = 110/172 (63%), Positives = 136/172 (79%), Frame = +2	TBLASTN Score = 560 (202.2 bits), Expect = 2.6e-54, P = 2.6e-54
One EST contig from poplars	<i>Populus tremula</i> x <i>Populus tremuloides</i>	gil24055393 gil24048976	Length = 470 Identities = 81/107 (75%), Positives = 90/107 (84%), Frame = +1	TBLASTN Score = 396 (144.5 bits), Expect = 5.7e-37, P = 5.7e-37
One EST from rice	<i>Oryza sativa</i>	gil18388132	Length = 481 Identities = 23/50 (46%), Positives = 33/50 (66%), Frame = +1	TBLASTN Score = 94 (38.1 bits), Expect = 4.4e-07, Sum P(2) = 4.4e-07
One EST from tomato	<i>Lycopersicon esculentum</i>	gil6058991	Length = 615 Identities = 52/134 (38%), Positives = 84/134 (62%), Frame = +3	TBLASTN Score = 241 (89.9 bits), Expect = 1.8e-20, P = 1.8e-20
One EST contig from potato	<i>Solanum tuberosum</i>	gil21373039 gil21373040 gil17076330	Length = 491 Identities = 33/66 (50%), Positives = 43/66 (65%), Frame = +1	TBLASTN Score = 168 (64.2 bits), Expect = 7.4e-13, P = 7.4e-13
One EST contig from sugar beet	<i>Beta vulgaris</i>	gil26112020 gil15427905 gil26119066 gil26118797 gil26118970 gil26118882	Length = 1155 Identities = 40/142 (28%), Positives = 74/142 (52%), Frame = +1	TBLASTN Score = 127 (49.8 bits), Expect = 9.9e-07, P = 9.9e-07

Closest Arabidopsis homologs:

At1g47750	<i>Arabidopsis thaliana</i>	>gil18401872 >gil25405892 >gil9802590 	Length = 248 Identities = 233/248 (93%), Positives = 233/248 (93%)	BLASTP Score = 1166 (415.5 bits), Expect = 8.0e-118, P = 8.0e-118
At3g47430	<i>Arabidopsis thaliana</i>	>gil15232797 >gil7487306 >gil5541710 >gil25083111 >gil30023692 	Length = 227 Identities = 58/153 (37%), Positives = 89/153 (58%)	BLASTP Score = 229 (85.7 bits), Expect = 1.6e-18, P = 1.6e-18

At2g45740	<i>Arabidopsis thaliana</i>	>gi 18406877 >gi 30690116 >gi 15450880 >gi 20197204 >gi 21537163	Length = 236 Identities = 52/227 (22%), Positives = 99/227 (43%)	BLASTP Score = 132 (51.5 bits), Expect = 2.7e-06, P = 2.7e-06
At3g61070	<i>Arabidopsis thaliana</i>	>gi 30695340 >gi 27754657 >gi 28394051	Length = 231 Identities = 38/142 (26%), Positives = 71/142 (50%)	BLASTP Score = 120 (47.3 bits), Expect = 7.9e-05, P = 7.9e-05
At1g01820	<i>Arabidopsis thaliana</i>	>gi 18378887 >gi 25511713 >gi 8671852 >gi 12083290 >gi 17381255 >gi 20453367 >gi 21555588	Length = 235 Identities = 37/142 (26%), Positives = 75/142 (52%)	BLASTP Score = 118 (46.6 bits), Expect = 0.00015, P = 0.00015

At1g47750 is a non-secretory protein (SignalP) that has one transmembrane domain (transmembrane helix = aa residues 98-120; predicted by TMHMM). The cellular localization of At1g47750 is predicted by Psort2 (32.0 %: cytoplasmic, 32%: nuclear, 12%: mitochondrial, 8%: vesicles of secretory system, 8%: cytoskeletal). In summary, At1g47750 is likely to be a transmembrane protein.

Pfam analysis showed that At1g47750 is a member of the peroxisomal biogenesis factor 11 (PEX11) family of proteins (PF05648). This family consists of several peroxisomal biogenesis factor 11 (PEX11) proteins from several eukaryotic species (e.g. *Arabidopsis*, yeast, mouse and human). In yeast, mouse and human, overexpression of PEX11 promotes peroxisome division. By contrast, loss of PEX11 results in reduced peroxisome abundance (see Li and Gould 2003; Li et al 2002; Carlo et al, 2000).

Since peroxisome is required for fatty acid metabolism (e.g. beta oxidation of fatty acids). One could envision that altered expression of At1g47750 can lead to change in seed oil content in *Arabidopsis*.

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
PF05648	1/1	18	245 ..	1	263 []	321.3	1.5e-93

EXAMPLE 4Generation of Mutated Plants with a HIO110.3 Phenotype

The invention further provides a method of identifying plants that have mutations in, or an allele of, endogenous HIO110.3 that confer a HIO110.3 phenotype, and
5 generating progeny of these plants that also have the HIO110.3 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. HIO110.3-specific PCR is used to identify whether a
10 mutated plant has a HIO110.3 mutation. Plants having HIO110.3 mutations may then be tested for the HIO110.3 phenotype, or alternatively, plants may be tested for the HIO110.3 phenotype, and then HIO110.3-specific PCR is used to determine whether a plant having the HIO110.3 phenotype has a mutated HIO110.3 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 HIO110.3 gene or orthologs of HIO110.3 that may confer the HIO110.3 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 HIO110.3 nucleic acid is used to identify whether a plant having a HIO110.3 phenotype has a mutation in endogenous
25 HIO110.3 or has a particular allele that causes the HIO110.3 phenotype compared to plants lacking the mutation or allele, and generating progeny of the identified plant that have inherited the HIO110.3 mutation or allele and have the HIO110.3 phenotype.

EXAMPLE 5Recapitulation of the High Oil Phenotype

To confirm that over-expression of At1g47750 causes the high seed oil phenotype in HIO110.3, Arabidopsis plants of the Col-0 ecotype were transformed by Agrobacterium mediated transformation with a construct containing the coding sequences of the HIO110.3

gene (At1g47750) behind the CsVMV promoter and in front of the nos terminator or a control gene unrelated to pathogen resistance. Both of these constructs contain the *nptII* gene to confer kanamycin resistance in plants. T1 seed was harvested from the transformed plants and transformants selected by germinating seed on agar medium
5 containing kanamycin. Kanamycin resistant transformants were transplanted to soil after 7 days. Control plants were germinated on agar medium without kanamycin, transplanted to soil after 7 days. Twenty plants containing the CsVMV::HIO110.3 transgene along with 10 Col-0 control plants were grown in the same flat in the growth room. All plants were allowed to self-fertilize and set seed.

10 To evaluate the high seed oil phenotype T2 seed pools were analyzed by Near Infrared Spectroscopy (NIR) 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 (ren oil 1473 1d +
15 sline.q2, Predicts Hexane Extracted Oil), followed the general method of AOCS Procedure AM1-92, Official Methods and Recommended Practices of the American Oil Chemists Society, 5th Ed., AOCS, Champaign, Ill. Seed from transgenic plants and control plants grown in the same flat were compared. T2 seed from plants containing the CsVMV::HIO110.3 transgene contained an average of 3% more oil than the seed from the
20 control plants (Table 2). This increase was determined to be statistically significant ($P < 0.05$) by a T-test.

TABLE 2

Plant ID	Transgene	Predicted average	Relative value average
Z004086001	CsVMV::HIO110.3	30.8579	96.4578
Z004086002	CsVMV::HIO110.3	31.2906	97.8105
Z004086003	CsVMV::HIO110.3	33.1408	103.5938
Z004086004	CsVMV::HIO110.3	34.0951	106.577
Z004086005	CsVMV::HIO110.3	33.1637	103.6655
Z004086006	CsVMV::HIO110.3	33.8006	105.6564
Z004086007	CsVMV::HIO110.3	33.1451	103.6073
Z004086008	CsVMV::HIO110.3	31.9286	99.8047
Z004086009	CsVMV::HIO110.3	31.8948	99.6991
Z004086010	CsVMV::HIO110.3	33.1731	103.6949
Z004086011	CsVMV::HIO110.3	33.186	103.7352
Z004086012	CsVMV::HIO110.3	31.9956	100.0143
Z004086013	CsVMV::HIO110.3	34.8563	108.9563
Z004086014	CsVMV::HIO110.3	32.504	101.6036
Z004086017	CsVMV::HIO110.3	32.1516	100.5018
Z004086018	CsVMV::HIO110.3	33.3858	104.3597
Z004086020	CsVMV::HIO110.3	33.0029	103.163
Z004086021	CsVMV::HIO110.3	33.9408	106.0946
Z004086022	CsVMV::HIO110.3	34.4272	107.615
Z004074001	None	32.3752	101.2007
Z004074002	None	31.4535	98.3197
Z004074003	None	31.0313	96.9998
Z004074004	None	30.1941	94.383
Z004074005	None	31.5536	98.6324
Z004074006	None	30.8081	96.3021
Z004074007	None	33.3079	104.1162
Z004074008	None	32.7195	102.277
Z004074009	None	32.4597	101.4649
Z004074010	None	34.0078	106.3041

REFERENCES

- 5 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.
- 10 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.
- 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.
- 5 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.
- 10 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.
- 15 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.
- 20 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.
- 25 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.
- Harlow E and Lane D, Antibodies: A Laboratory Manual, Cold Spring Harbor
- 30 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.

- 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.
- 5 Kaneko, T. *et al.*, DNA Res. 7:217-221, 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.
- Kline *et al.*, Nature (1987) 327:70.
- 10 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.
- Lemieux B., *et al.*, 1990, Theor Appl Genet 80, 234-240.
- Ludwig, A., *et al.*, Plant J. 24:503-509, 2000.
- 15 Meyer, S., *et al.*, Plant J. 24:869, 2000.
- Moreno, M., *et al.*, Yeast 10(11):1447-1457, 1994.
- 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.
- 20 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.
- Robinson, M.S., J. Cell. Biol. 108(3):833-842, 1989.
- Sambrook *et al.*, Molecular Cloning: A Laboratory Manual (Second Edition), Cold
- 25 Sasaki, T., Nature 420:312-316, 2002.
- Sakai, Y., *et al.*, J. Bacteriol. 177(23):6773-6781, 1995.
- 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.
- 30 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.
- Theologis, A. *et al.*, Nature 408:816-820, 2000.
- Thompson JD *et al.*, Nucleic Acids Res 22:4673-4680, 1994.
- Tucker, K.L. *et al.*, Nucleic Acids Res 18(17):5306, 1990.
- 5 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.
- Voglmaier, S. M. *et al.*, Biochem. Biophys. Res. Commun. 187(1):L158-163, 1992.
- 10 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.
- 15 Weise A, *et al.*, Plant Cell 12:1345-1355, 2000.
- Wilson K *et al.*, Plant Cell 8: 659-671, 1996.
- Yadav NS *et al.*, (1993) Plant Physiol 103, 467-476.
- Li X, Gould SJ. (2003) J Biol Chem. 278, 17012-17020.
- Li X, *et al.*, (2002) Mol Cell Biol 22,8226-8240.
- 20 van Roermund CW, *et al.*, (2000) J Cell Biol. 150,489-498.

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 HIO110.3 polypeptide, whereby the transgenic plant has a high oil phenotype.
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:
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 HIO110.3 polypeptide, and
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.
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. The plant of claim 7, wherein the plant is selected from the group consisting of a plant grown from said progenitor cells, a plant that is the direct progeny of a plant grown from

said progenitor cells, and a plant that is the indirect progeny of a plant grown from said progenitor cells.

10. A method of generating a plant having a HIO110.3 phenotype comprising identifying
5 a plant that has an allele or mutation in its HIO110.3 gene that results in a high oil
phenotype compared to plants lacking the allele or mutation and generating progeny of
said identified plant, wherein the generated progeny inherit the allele and have the
HIO110.3 phenotype.
- 10 12. The method of claim 10 that employs candidate gene/QTL methodology.
13. The method of claim 10 that employs TILLING methodology.

SEQUENCE LISTING

<110> Agrinomics LLC

<120> Generation of Plants with Altered Oil Content

<130> AG04-036C-PC

<150> 60/487,380

<151> 2003-07-14

<160> 9

<170> PatentIn version 3.2

<210> 1

<211> 747

<212> DNA

<213> Arabidopsis thaliana

<400> 1

```

atggctacga aagctccaga gaaaatcacc aaaccaaag acagagattt cctaaatcat      60
ctcgaaacat atctctccaa aagagatggt gtagacaaac tcctgaaaat ctctcgttac      120
gcaacaaaaa tcattcctgc atcatcacta atcccagaaa ccagatccat aatccccga      180
ctcaaaagct tcgaatcaag cgtcggagtc agtcgcaaag ccttcctgtct cggaaaattc      240
gtacaagaca tcaacgcatt aagatcctca cgatgggatt caaatcacga gctcgtctta      300
ctcataatcg cgtacgggtg tgaaggatta tactatttcg ttgaacaatt catatggtta      360
acgaaatcag gactgatcga tgcgaaacat tcgaaatggc ttcagaagat tagtgcttgg      420
gctgaattgg ttggttatgt aggaagtgtt tcgattaaga ttagggattt gaggaaattg      480
aatgatgaag agagttgtgt tgcttcgacg atcgagatct cggtttcgag aggattagct      540
tgtgatggtg aagatgagaa gatgaagatg attaaggaga agaagacatt gaaggtttta      600
tctattttgc aggatcttgc tgatggtttg atgacgattg ctgatattcg tgatggtaag      660
ggagtgttat cggctcctaa tgtgatatct tctgcagggt tgttttctgc tattgttagt      720
actcataaga actggatctc ttgttga                                          747

```

<210> 2

<211> 248

<212> PRT

<213> Arabidopsis thaliana

<400> 2

```

Met Ala Thr Lys Ala Pro Glu Lys Ile Thr Lys Pro Lys Asp Arg Asp
1           5           10           15

```

```

Phe Leu Asn His Leu Glu Thr Tyr Leu Ser Lys Arg Asp Gly Val Asp
          20           25           30

```

Lys Leu Leu Lys Ile Ser Arg Tyr Ala Thr Lys Ile Ile Leu Ala Ser
 35 40 45
 Ser Leu Ile Pro Glu Thr Arg Ser Ile Ile Pro Arg Leu Lys Ser Phe
 50 55 60
 Glu Ser Ser Val Gly Val Ser Arg Lys Ala Phe Arg Leu Gly Lys Phe
 65 70 75 80
 Val Gln Asp Ile Asn Ala Leu Arg Ser Ser Arg Trp Asp Ser Asn His
 85 90 95
 Glu Leu Val Leu Leu Ile Ile Ala Tyr Gly Gly Glu Gly Leu Tyr Tyr
 100 105 110
 Phe Val Glu Gln Phe Ile Trp Leu Thr Lys Ser Gly Leu Ile Asp Ala
 115 120 125
 Lys His Ser Lys Trp Leu Gln Lys Ile Ser Ala Trp Ala Glu Leu Val
 130 135 140
 Gly Tyr Val Gly Ser Val Ser Ile Lys Ile Arg Asp Leu Arg Lys Leu
 145 150 155 160
 Asn Asp Glu Glu Ser Cys Val Ala Ser Thr Ile Glu Ile Ser Val Ser
 165 170 175
 Arg Gly Leu Ala Cys Asp Gly Glu Asp Glu Lys Met Lys Met Ile Lys
 180 185 190
 Glu Lys Lys Thr Leu Lys Val Leu Ser Ile Leu Gln Asp Leu Ala Asp
 195 200 205
 Gly Leu Met Thr Ile Ala Asp Ile Arg Asp Gly Lys Gly Val Leu Ser
 210 215 220
 Ala Pro Asn Val Ile Ser Ser Ala Gly Leu Phe Ser Ala Ile Val Ser
 225 230 235 240
 Thr His Lys Asn Trp Ile Ser Cys
 245

<210> 3
 <211> 602
 <212> DNA
 <213> Arabidopsis thaliana

```

<400> 3
tcgggcacgag ggctccccac caccgacgag ctctctccct cttccctcgc cccagatgga      60

gtagcagggc cggatcctgc cgccatccga ggctcgcaag accgccgccg cagcgcctcc      120

gcctccccgc gacttcctcg cctacctga ggcctatctg gcccgccacg acggcgctga      180

caagctcctt aagatctcac actacgctac gcgcctcgcc ctgggccctt cgctgcaggg      240

cctctgcccc gcaggcctcc gcgcgcctca agtccttcga gtccagcctc ggctcagac      300

gcaaggcctt ccgcctcagc aagttcgtcc agacgtcaac gccctccgcg cgcactccgg      360

ccccctccct ctgcccttcg tgctcctcgc ctacggcggc gagggcgctt actacatcat      420

cgggcagttc gtctggctcg cgaaagctgc cctcctcccg gcgcagctcc tccccgcct      480

tcagcgcctc agcgcttggg ccgagctgct gggttacgtc ggctccaaca ccatcaagct      540

ggaagtgggtg atgaagatgg agtcttcaat caagatgcgg ctggcggagg gttgtgggga      600

gg                                                                                   602

```

```

<210> 4
<211> 697
<212> DNA
<213> Arabidopsis thaliana

```

```

<220>
<221> misc_feature
<222> (633)..(633)
<223> n is a, c, g, or t

```

```

<400> 4
ggcttatttt ttttcctctt tcgaaaagat tggaccaaac aagcctttat cagaagatta      60

gagaatcgaa aaacgacaca ttgagctgac gctggagagg gcgtgacacc cagagaccct      120

aatctgcttg cttgccttaa aaaaaacgaa tttttaatta aaaaaggaga aaagtttggtg      180

gatttcgaga taaaggatga gtacactgga tgcaactaga gcagaacttg ctctttagt      240

tttgtatttg aataaggctg aagcaaggga taagatatgc agggctatac aatatggctc      300

gaaattccta agtgatgggc agcctggcac agcacaaaat gttgacaaat caaccagctt      360

ggcacggaaa gttttccgtc tttttaagtt tgtgaacgat ttgcatgggc ttattagccc      420

agttcctcaa ggaactcccc ttccacttgt tctgctagga aagtcaaaga atgcactggt      480

gtcaactttc ttatttctcg atcaaattgt gtggcttggc agaacaggca tctataagaa      540

caaagaacgt gctgagttga ttggacgaat atcccttttc tgttgatgg gttcttcaat      600

ctgtaccacc ttggttgagc ttggggagct canaaggcta tctaaatcaa tgaaaaagtt      660

ggaaaaaaat cttaggaca gtgataagca ccagaac                                         697

```

```

<210> 5

```

<211> 572
 <212> DNA
 <213> Arabidopsis thaliana

<400> 5
 cggttgctgtc gacgccctcc gcgccacccc gcacctccc cgggccctcg cgctactcgc 60
 ctacggcggg gagggcgctct actacttcgt cgagcagttc gtctggatcg ccaaggccgg 120
 gctcctgcct gcgcaccttc ttccgcgcct ccagcgtctc agcgccctggg ctgagctgct 180
 gggctacgtc ggttccattt taattaatct ggaggagggtg gctaagatcg aatcttcgat 240
 aaagaagcgg ctctgtgaag actgtgataa ggagagcgag gcggtgaaga ctatgcaggg 300
 taaactgctg ctgaagcggg tgtcagtggt gcaggacgtg gcggatgccg tcatggctct 360
 gggagacgtc accgatggca agggtttgct cgggagctcc acgctgatgg cagcagctgg 420
 actgctctcg gcgttgatca gcacgcataa gaattggaat tcttggtgaa gcctggaggt 480
 acatgttata gatagcacat catataatga aataaagctc tgattatagc gataaccaag 540
 agtttcctct ggaaatatgg atgctcctaa aa 572

<210> 6
 <211> 862
 <212> DNA
 <213> Arabidopsis thaliana

<220>
 <221> misc_feature
 <222> (725)..(725)
 <223> n is a, c, g, or t

<220>
 <221> misc_feature
 <222> (732)..(732)
 <223> n is a, c, g, or t

<220>
 <221> misc_feature
 <222> (735)..(735)
 <223> n is a, c, g, or t

<220>
 <221> misc_feature
 <222> (738)..(738)
 <223> n is a, c, g, or t

<220>
 <221> misc_feature
 <222> (743)..(744)
 <223> n is a, c, g, or t

<220>
 <221> misc_feature
 <222> (746)..(749)
 <223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (753)..(753)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (756)..(756)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (764)..(766)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (769)..(769)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (772)..(774)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (778)..(781)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (783)..(783)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (790)..(790)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (793)..(793)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (800)..(800)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (834)..(834)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (845)..(845)
<223> n is a, c, g, or t

<400> 6
atcatttctt tgtcatgtct tcctttcttt cttttgtaga cacatctaac atttttcttt

60

```

agaagacttg cagttgcagc agcatctgtt cccagtcggt ggcagaacag caccaccctc 120
aatggattcc aattccaaac cttccgcacc acccccgaac cctaacccta accctaattc 180
atcacccctcg gagaaacacc agagagacct actgaaccac gtcgaagcct acctcgcgaa 240
gcgcgacggc gtggacaagc tcctcaagat ttccgcgtac accgcgaaac tcattcctcg 300
cacttcgctc ctccaatcca acaaccctac cctctccac cgctaaaaat ccttcgaatc 360
cagcgtcggc gtgagccgaa aggccttcag actcggtaaa ttcgttcaag acctaaacgc 420
cctcaaaacc acccatcccc attogaacct cgatttgttc ctctccgtcc tcgcctacgg 480
aggcgaaggc ttgtactact tcgtggagca gttcgtcttg ctgcgaaat cggggctgat 540
tgatgccaaag cactcgcgaa agttccagaa attgagcgcg tgggcggagt tcgtgggtta 600
cttcgggagc gtggcactca aactcgggga tttgggtgcg atagcccagg aggaagcgtg 660
tttgaagtcg agcgtggaga tcacgcgcct gagagagggg ggggtgtgtt atgaggaggc 720
gcagntgggg angngngng gcnnngnnnc ggngangagg gttnnngng gnnnggannn 780
ngncgatgcn ggnatgggt tagatgatgt tttggatgga aatggaccgt tttngaaacc 840
ggttntgatg gcttccccgg gg 862

```

```

<210> 7
<211> 470
<212> DNA
<213> Arabidopsis thaliana

```

```

<400> 7
tctcccctcg catctccctc tctctcgca aagacttcta ccgactaaga aattcctgct 60
acacactctc accaggaaat ttctctctg ttttgacca ccaccaccac gacgatggat 120
cccaagcctt caccaaacc acaaaacgcc aatgaaatcc ctaataaaca gggacaaaga 180
gactttttga atcatttaga ggtatacctt gcaaagcgag atgggggtga caagctctc 240
aagatctcac gttacgcaac caagatcatc ttagcctcat ccgtcctccc cgaaacctta 300
atcgtaacga aacgactaaa gagtttcgaa tcaagtgttg gactcagtcg caaagcattc 360
cgacttgga aattcgttca agatatcaac gcattaagag gctctcattt tgattcaaaa 420
caagaaatca tcctttccat cattgcctat ggaggtaaag gcttgtatta 470

```

```

<210> 8
<211> 491
<212> DNA
<213> Arabidopsis thaliana

```

```

<400> 8
ggcacgaggg agaaattgag gaaattgaga ttgaagaagt tgatgaagag gttatcagtg 60
attcaagatt ttgtgatgg gttaatggct ttagctgata ttagtgatgg taaagggatg 120

```

```

ctttctgcac cattgttggt gtcacagca gggttgttgt cagctcttat tagcacacat 180
aagaattgga tctcttgctg atcatccata gctggatttg gaggtatggt ttgtgcttaa 240
tcaaattgga gtttacgttg tcgtaggatt agttagtgat aaagattcat tttttaaaag 300
taatgtaatg tgttctcccc aaaataagaa gggaattggg agaaggtata caagaggtgt 360
gaagttatct tattgcagat agaagatttc tatctcaaag atagttactt tagttcctca 420
tgacaatttt acttggtgca tcatgaaaga ttctggcttt cgatagttcg taaagattta 480
ccatttcagc t 491

```

```

<210> 9
<211> 1155
<212> DNA
<213> Arabidopsis thaliana

```

```

<400> 9
ggagagtgga gagattatcc aaactccaaa ttgtttgggt agtaagagga gaagagagga 60
ccaaacaagc ccttatcaaa ttttagcgga agagaaaaaa ctctttcttc aatactgacg 120
caattgataa attctcagga caaataaatt atctccccct aagctgctgg aggagagaga 180
aatcaacac agaagtttca atttcttagc ttttaaggac agtagagaag atataatatg 240
agtaactag atgtgccag ggcagaactt gcccttgagc ttctgtatct gaacaaagca 300
gaagctaggg acaagatatg cagggcaatc cagtatgggt ccaagtttct gagtgatgga 360
cagccaggta cagcacaaaa tgttgataaa tcaaccagtc tggcaagaaa agtcttccgt 420
ctctttaagt ttgttaatga tttgcatgct cttattagcc cagttcctgc tggaactcct 480
cttcccttaa ctttacttgg aaagtccaaa aatgcactat tgtccacatt tttgttcctt 540
gatcaatttg tctggcttgg tagaagtggc atctataaga acaaggagcg aactgacctg 600
attgccaaaga tatctcttta ctgttggatg agttcctctg tctgtacttc tttggttgag 660
cttggggaac ttggaaggct ttctgcttca atgaagaagc tggagaagga gcttaagggc 720
agtgggaaac acgacaatga tcaataccaa gccagggttc agaagtataa tgacaggtca 780
ttggcccttag ttaaagccgc cttggatata gttgtggctg cggggctgct tcaattggca 840
cccaagaaaa ttactcctcg tgtaactggg gcctttgggt ttgctacatc tctcatctca 900
tgttatcagt tacttccggc acggccaaag tccaaaacat cttgaacaag ggaaaacagc 960
ttcctttaag cccctgtgga ttataaaaag ccatgatcct tgttgtgtat gtttgaaaga 1020
atctagaaga attttagtag ggtgacattt tcacctttaa ttaagaacca ttcttgtatt 1080
tcctttcatg ccttgtgtta acagttcacg atagcaataa tatgtttgtt gctcttgttt 1140
aaaaaaaaa aaaaa 1155

```