



(86) Date de dépôt PCT/PCT Filing Date: 2007/11/14
(87) Date publication PCT/PCT Publication Date: 2008/05/22
(45) Date de délivrance/Issue Date: 2016/03/01
(85) Entrée phase nationale/National Entry: 2009/05/14
(86) N° demande PCT/PCT Application No.: US 2007/084709
(87) N° publication PCT/PCT Publication No.: 2008/061153
(30) Priorité/Priority: 2006/11/15 (US60/866,053)

(51) Cl.Int./Int.Cl. *C12N 15/82* (2006.01),
A01H 5/00 (2006.01)
(72) Inventeurs/Inventors:
DAVIES, JOHN P., US;
NG, HEIN TSOENG (MEDARD), US;
WAGNER, D. RY, US
(73) Propriétaire/Owner:
AGRIGENETICS, INC., US
(74) Agent: GOUDREAU GAGE DUBUC

(54) Titre : GENERATION DE PLANTES EXPRIMANT UN POLYPEPTIDE IMQ ET COMPORTANT UNE TENEUR EN
PROTEINES, FIBRES OU HUILE MODIFIEE
(54) Title: GENERATION OF PLANTS EXPRESSING AN IMQ POLYPEPTIDE AND HAVING ALTERED PROTEIN,
FIBER OR OIL CONTENT

(57) **Abrégé/Abstract:**

The present invention is directed to plants that display an improved oil quantity phenotype or an improved meal quality phenotype due to altered expression of an IMQ nucleic acid, such as a nucleic acid that encodes or is complementary to a sequence that encodes an IMQ polypeptide comprising an amino acid sequence at least 80% identical to the amino acid sequence as set forth in SEQ ID NO: 80. The invention is further directed to methods of generating plants with an improved oil quantity phenotype or improved meal quality phenotype.

ABSTRACT

The present invention is directed to plants that display an improved oil quantity phenotype or an improved meal quality phenotype due to altered expression of an IMQ nucleic acid, such as a nucleic acid that encodes or is complementary to a sequence that encodes an IMQ polypeptide comprising an amino acid sequence at least 80% identical to the amino acid sequence as set forth in SEQ ID NO: 80. The invention is further directed to methods of generating plants with an improved oil quantity phenotype or improved meal quality phenotype.

GENERATION OF PLANTS EXPRESSING AN IMQ POLYPEPTIDE AND HAVING ALTERED PROTEIN, FIBER OR OIL CONTENT

5

10

FIELD OF THE DISCLOSURE

The present disclosure is related to transgenic plants with altered oil, protein, and/or fiber content, as well as methods of making plants having altered oil, protein, and/or fiber content and producing oil from such plants.

15

BACKGROUND

The ability to manipulate the composition of crop seeds, particularly the content and composition of seed oil and protein, as well as the available metabolizable energy (“AME”) in the seed meal in livestock, has important applications in the agricultural industries, relating both to processed food oils and to animal feeds. 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 U.S. soybean crop is crushed by the soy processing industry. Soy processing yields purified oil, which is sold at high value, while the remaining seed meal is sold for livestock feed (U.S. Soybean Board, 2001 Soy Stats). Canola seed is also crushed to produce oil and the co-product canola meal (Canola Council of Canada). Canola meal contains a high percentage of protein and a good balance of amino acids but because it has a high fiber and phytate content, it is not readily digested by livestock (Slominski, B.A., *et al.*, 1999 Proceedings of the 10th International Rapeseed Congress, Canberra, Australia) and has a lower value than soybean meal.

30

Over 55% of the corn produced in the U.S. is used as animal feed (Iowa Corn Growers Association). The value of the corn is directly related to its ability to be

digested by livestock. Thus, it is desirable to maximize both oil content of seeds and the AME of meal. For processed oilseeds such as soy and canola, increasing the absolute oil content of the seed will increase the value of such grains, while increasing the AME of meal will increase its value. For processed corn, either an
5 increase or a decrease in oil content may be desired, depending on how the other major constituents are to be used. Decreasing oil may improve the quality of isolated starch by reducing undesired flavors associated with oil oxidation. Alternatively, when the starch is used for ethanol production, where flavor is unimportant, increasing oil content may increase overall value.

10 In many feed grains, such as corn and wheat, it is desirable to increase seed oil content, because oil has higher energy content than other seed constituents such as carbohydrate. Oilseed processing, like most grain processing businesses, is a capital-intensive business; thus small shifts in the distribution of products from the low valued components to the high value oil component can have substantial
15 economic impacts for grain processors. In addition, increasing the AME of meal by adjusting seed protein and fiber content and composition, without decreasing seed oil content, can increase the value of animal feed.

Biotechnological manipulation of oils has been shown to provide compositional alteration and improvement of oil yield. Compositional alterations
20 include high oleic acid soybean and corn oil (U.S. Patent Nos. 6,229,033 and 6,248,939), and laurate-containing seeds (U.S. Patent 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
25 biotechnology in this area is High-Oil Corn (HOC) technology (DuPont, U.S. Patent 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
30 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 content in current HOC fields has 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
5 these reasons, an alternative high oil strategy, particularly one that derives from an altered biochemical output, would be especially valuable.

Manipulation of seed composition has identified several components that improve the nutritive quality, digestibility, and AME in seed meal. Increasing the lysine content in canola and soybean (Falco *et al.*, 1995 *Bio/Technology* 13:577-582)
10 increases the availability of this essential amino acid and decreases the need for nutritional supplements. Soybean varieties with increased seed protein were shown to contain considerably more metabolizable energy than conventional varieties (Edwards *et al.*, 1999, *Poultry Sci.* 79:525-527). Decreasing the phytate content of corn seed has been shown to increase the bioavailability of amino acids in animal
15 feeds (Douglas *et al.*, 2000, *Poultry Sci.* 79:1586-1591) and decreasing oligosaccharide content in soybean meal increases the metabolizable energy in the meal (Parsons *et al.*, 2000, *Poultry Sci.* 79:1127-1131).

Soybean and canola are the most obvious target crops for the processed oil and seed meal markets since both crops are crushed for oil and the remaining meal
20 sold for animal feed. A large body of commercial work (*e.g.*, U.S. Patent No. 5,952,544; PCT Application No. 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.
25 For instance, screens using chemically mutagenized populations have identified lipid mutants whose seeds display altered fatty acid composition (Lemieux *et al.*, 1990, *Theor. Appl. Genet.* 80, 234-240; James and Dooner, 1990, *Theor. Appl. Genet.* 80, 241-245). T-DNA mutagenesis screens (Feldmann *et al.*, 1989, *Science* 243: 1351-1354) that detected altered fatty acid composition identified the omega 3 desaturase
30 (*FAD3*) and delta-12 desaturase (*FAD2*) genes (U.S. Pat No 5952544; Yadav *et al.*, 1993, *Plant Physiol.* 103, 467-476; Okuley *et al.*, 1994, *Plant Cell* 6(1):147-158). 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, *Plant Physiol.* 118:91-101).

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
5 accumulation in seeds (Katavic V *et al.*, 1995, *Plant Physiol.* 108(1):399-409). It was further shown that seed-specific over-expression of the DGAT cDNA was associated with increased seed oil content (Jako *et al.*, 2001, *Plant Physiol.* 126(2):861-74). *Arabidopsis* is also a model for understanding the
10 accumulation of seed components that affect meal quality. For example, *Arabidopsis* contains albumin and globulin seed storage proteins found in many dicotyledonous plants including canola and soybean (Shewry 1995, *Plant Cell* 7:945-956). The biochemical pathways for synthesizing components of fiber, such as cellulose and lignin, are conserved within the vascular plants, and mutants of
15 *Arabidopsis* affecting these components have been isolated (reviewed in Chapel and Carpita 1998, *Current Opinion in Plant Biology* 1:179-185).

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
20 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, *Science* 258: 1350-1353; Weigel D *et al.*, 2000, *Plant Physiology*, 122:1003-1013). The inserted construct provides a molecular tag for rapid identification of the native plant whose mis-
25 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
30 associated with these phenotypes (Wilson *et al.*, 1996, *Plant Cell* 8: 659-671; Schaffer *et al.*, 1998, *Cell* 93: 1219-1229; Fridborg *et al.*, 1999, *Plant Cell* 11: 1019-1032; Kardailsky *et al.*, 1999, *Science* 286: 1962-1965; and Christensen S *et al.*,

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

SUMMARY

5 Provided herein are transgenic plants having an Improved Seed Quality phenotype. Transgenic plants with an Improved Seed Quality phenotype may include an improved oil quantity and/or an improved meal quality. Transgenic plants with improved meal quality have an Improved Meal Quality (IMQ) phenotype and transgenic plants with improved oil quantity have an Improved Oil Quantity (IOQ) 10 phenotype. The IMQ phenotype in a transgenic plant may include altered protein and/or fiber content in any part of the transgenic plant, for example in the seeds. The IOQ phenotype in a transgenic plant may include altered oil content in any part of the transgenic plant, for example in the seeds. In particular embodiments, a transgenic plant may include an IOQ phenotype and/or an IMQ phenotype. In some 15 embodiments of a transgenic plant, the IMQ phenotype may be an increase in protein content in the seed and/or a decrease in the fiber content of the seed. In other embodiments of a transgenic plant, the IOQ phenotype is an increase in the oil content of the seed (a high oil phenotype). Also provided is seed meal derived from the seeds of transgenic plants, wherein the seeds have altered protein content and/or altered 20 fiber content. Further provided is oil derived from the seeds of transgenic plants, wherein the seeds have altered oil content. Any of these changes can lead to an increase in the AME from the seed or seed meal from transgenic plants, relative to control, non-transgenic, or wild-type plants. Also provided herein is meal, feed, or food produced from any part of the transgenic plant with an IMQ phenotype and/or 25 IOQ phenotype.

 The present invention also provides a transgenic plant cell, comprising a plant transformation vector comprising a nucleotide sequence that encodes or is complementary to a sequence that encodes an IMQ polypeptide comprising an amino acid sequence at least 80% identical to the amino acid sequence as set forth in SEQ ID 30 NO: 80, whereby the transgenic plant cell is of a transgenic plant that has an improved seed quality phenotype, relative to control plants, wherein the improved seed quality phenotype comprises an increase in oil content.

The present invention also provides a method of producing an improved seed quality phenotype in a plant, said method comprising:

5 a) introducing into progenitor cells of the plant a plant transformation vector comprising a nucleotide sequence that encodes an IMQ polypeptide comprising an amino acid sequence at least 80% identical to the amino acid sequence set forth in SEQ ID NO: 80, and

10 b) growing the transformed progenitor cells to produce a transgenic plant, wherein the nucleotide sequence is expressed, and the transgenic plant exhibits an improved seed quality phenotype relative to control plants, and wherein the improved seed quality phenotype comprises an increase in oil content, thereby producing the improved seed quality phenotype in the plant.

The present invention also provides a feed, meal, grain or food comprising a polypeptide encoded by a nucleic acid sequence at least 80% identical to the nucleic acid sequence as set forth in SEQ ID NO: 79.

15 The present invention also provides a feed, meal, grain or food comprising a polypeptide comprising an amino acid sequence at least 80% identical to the amino acid sequence as set forth in SEQ ID NO: 80.

The present invention also provides a method of producing an improved seed quality phenotype in a plant, said method comprising:

20 a) introducing into progenitor cells of the plant a plant transformation vector comprising a nucleotide sequence that encodes an IMQ polypeptide comprising an amino acid sequence at least 95% identical to the full length of the amino acid sequence set forth in SEQ ID NO: 80,

25 b) growing the transformed progenitor cells to produce a transgenic plant, wherein the nucleotide sequence is expressed, and the transgenic plant exhibits an improved seed quality phenotype relative to control plants, and wherein the improved seed quality phenotype comprises an increase in oil content, and

30 c) measuring the improved seed quality phenotype in the transgenic plant, relative to control plants, thereby producing the improved seed quality phenotype in the plant.

The present invention also provides a method of generating a plant having an improved seed quality phenotype comprising

identifying a plant that has an allele in an IMQ gene, where the wild-type IMQ gene has the nucleic acid sequence set forth as SEQ ID NO: 79, which allele results in improved seed quality phenotype compared to a control plant lacking the allele;

generating progeny of said identified plant, wherein the generated progeny
5 inherit the allele and have the improved seed quality phenotype; and

measuring the improved seed quality phenotype in the progeny comprising the allele, relative to the control plant lacking the allele,
thereby generating a plant having an improved seed quality phenotype.

In certain embodiments, the disclosed transgenic plants comprise a
10 transformation vector comprising an IMQ nucleotide sequence that encodes or is complementary to a sequence that encodes an “IMQ” polypeptide. In particular embodiments, expression of an IMQ polypeptide in a transgenic plant causes an altered oil content, an altered protein content, and/or an altered fiber content in the transgenic plant. In preferred embodiments, the transgenic plant is selected from the
15 group consisting of plants of the *Brassica* species, including canola and rapeseed,

soy, corn, sunflower, cotton, cocoa, safflower, oil palm, coconut palm, flax, castor, peanut, wheat, oat and rice. Also provided is a method of producing oil or seed meal, comprising growing the transgenic plant and recovering oil and/or seed meal from said plant. The disclosure further provides feed, meal, grain, or seed
5 comprising a nucleic acid sequence that encodes an IMQ polypeptide. The disclosure also provides feed, meal, grain, or seed comprising the IMQ polypeptide, or an ortholog thereof.

Examples of the disclosed transgenic plant are produced by a method that comprises introducing into progenitor cells of the plant a plant transformation vector
10 comprising an IMQ nucleotide sequence that encodes, or is complementary to a sequence that encodes, an IMQ polypeptide, and growing the transformed progenitor cells to produce a transgenic plant, wherein the IMQ polynucleotide sequence is expressed, causing an IOQ phenotype and/or an IMQ phenotype in the transgenic plant. In some specific, non-limiting examples, the method produces transgenic
15 plants wherein expression of the IMQ polypeptide causes a high (increased) oil, high (increased) protein, and/or low (decreased) fiber phenotype in the transgenic plant, relative to control, non-transgenic, or wild-type plants.

Additional methods are disclosed herein of generating a plant having an IMQ and/or an IOQ phenotype, wherein a plant is identified that has an allele in its IMQ
20 nucleic acid sequence that results in an IMQ phenotype and/or an IOQ phenotype, compared to plants lacking the allele. The plant can generate progeny, wherein the progeny inherit the allele and have an IMQ phenotype and/or an IOQ phenotype. In some embodiments of the method, the method employs candidate gene/QTL methodology or TILLING methodology.

25 Also provided herein is a transgenic plant cell having an IMQ phenotype and/or an IOQ phenotype. The transgenic plant cell comprises a transformation vector comprising an IMQ nucleotide sequence that encodes or is complementary to a sequence that encodes an IMQ polypeptide. In preferred embodiments, the transgenic plant cell is selected from the group consisting of plants of the *Brassica*
30 species, including canola and rapeseed, soy, corn, sunflower, cotton, cocoa, safflower, oil palm, coconut palm, flax, castor, peanut, wheat, oat and rice. In other embodiments, the plant cell is a seed, pollen, propagule, or embryo cell. The

disclosure also provides plant cells from a plant that is the direct progeny or the indirect progeny of a plant grown from said progenitor cells.

DETAILED DESCRIPTION

5 Terms

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 disclosure. Practitioners are particularly directed to Sambrook *et al.* (*Molecular Cloning: A Laboratory Manual* (Second Edition), Cold Spring Harbor Press,
10 Plainview, N.Y.,1989) and Ausubel FM *et al.* (*Current Protocols in Molecular Biology*, John Wiley & Sons, New York, N.Y.,1993) for definitions and terms of the art. It is to be understood that this disclosure is not limited to the particular methodology, protocols, and reagents described, as these may vary.

As used herein, the term “IMQ phenotype” refers to plants, or any part of a
15 plant (for example, seeds, or meal produced from seeds), with an altered protein and/or fiber content (phenotype). As provided herein, altered protein and/or fiber content includes either an increased or decreased level of protein and/or fiber content in plants, seeds or seed meal. Any combination of these changes can lead to an IMQ phenotype. For example, in one specific non-limiting example, an IMQ
20 phenotype can refer to increased protein and decreased fiber content. In another specific non-limiting example, an IMQ phenotype can refer to unchanged protein and decreased fiber content. In yet another specific non-limiting example, an IMQ phenotype can refer to increased protein and unchanged fiber content. It is also provided that any combination of these changes can lead to an increase in the AME
25 (available metabolizable energy) from the seed or meal generated from the seed. An IMQ phenotype also includes an improved seed quality (ISQ) phenotype or an improved seed meal quality phenotype.

As used herein, the term “IOQ phenotype” refers to plants, or any part of a plant (for example, seeds), with an altered oil content (phenotype). As provided
30 herein, altered oil content includes an increased, for example a high, oil content in plants or seeds. In some embodiments, a transgenic plant can express both an IOQ phenotype and an IMQ phenotype. In specific, non-limiting examples, a transgenic

plant having a combination of an IOQ phenotype and an IMQ phenotype can lead to an increase in the AME (available metabolizable energy) from the seed or meal generated from the seed. An IOQ phenotype also includes an improved seed quality (ISQ) phenotype.

5 As used herein, the term “available metabolizable energy” (AME) refers to the amount of energy in the feed that is able to be extracted by digestion in an animal and is correlated with the amount of digestible protein and oil available in animal meal. AME is determined by estimating the amount of energy in the feed prior to feeding and measuring the amount of energy in the excreta of the animal
10 following consumption of the feed. In one specific, non-limiting example, a transgenic plant with an increase in AME includes transgenic plants with altered seed protein and/or fiber content and without a decrease in seed oil content (seed oil content remains unchanged or is increased), resulting in an increase in the value of animal feed derived from the seed.

15 As used herein, the term “content” refers to the type and relative amount of, for instance, a seed or seed meal component.

 As used herein, the term “fiber” refers to non-digestible components of the plant seed including cellular components such as cellulose, hemicellulose, pectin, lignin, and phenolics.

20 As used herein, the term “meal” refers to seed components remaining following the extraction of oil from the seed. Examples of components of meal include protein and fiber.

 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
25 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
30 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. Specific, non-limiting examples of a heterologous nucleic acid sequence include an IMQ nucleic acid sequence, or a fragment, derivative (variant), or ortholog thereof.

10 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 sequences.

15 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.

25 “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,

30

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
5 into a cell, includes “transfection,” “transformation,” and “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
10 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 from plant seeds, pollen, propagules, and embryos.

As used herein, the terms “native” and “wild-type” relative to a given plant
15 trait or phenotype refers to the form in which that trait or phenotype is found in the same variety of plant in nature. In one embodiment, a wild-type plant is also a control plant. In another embodiment, a wild-type plant is a non-transgenic plant.

As used herein, the term “modified” regarding a plant trait, refers to a change in the phenotype of a transgenic plant (for example, a transgenic plant with any
20 combination of an altered oil content, an altered protein content, and/or an altered fiber content) in any part of the transgenic plant, for example the seeds, relative to a similar non-transgenic plant. As used herein, the term “altered” refers to either an increase or a decrease of a plant trait or phenotype (for example, oil content, protein content, and/or fiber content) in a transgenic plant, relative to a similar non-
25 transgenic plant. In one specific, non-limiting example, a transgenic plant with a modified trait includes a plant with an increased oil content, increased protein content, and/or decreased fiber content relative to a similar non-transgenic plant. In another specific, non-limiting example, a transgenic plant with a modified trait includes unchanged oil content, increased protein content, and/or decreased fiber
30 content relative to a similar non-transgenic plant. In yet another specific, non-limiting example, a transgenic plant with a modified trait includes an increased oil content, increased protein content, and/or unchanged fiber content relative to a

similar non-transgenic plant. Specific, non-limiting examples of a change in phenotype include an IMQ phenotype or an IOQ phenotype.

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 plant (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 (for example, increased oil content, increased protein content, and/or decreased fiber content in seeds of 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 phenotype or quality. Such transgenic plants may have an improved phenotype, such as an IMQ phenotype or an IOQ phenotype.

The phrase “altered oil content phenotype” refers to a measurable phenotype of a genetically modified (transgenic) 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 (non-transgenic) plant. A high oil phenotype refers to an increase in overall oil content. The phrase “altered protein content phenotype” refers to measurable phenotype of a genetically modified plant, where the plant displays a statistically significant increase or decrease in overall protein content (i.e., the percentage of seed mass that is protein), as compared to the similar, but non-modified plant. A high protein phenotype refers to an increase in overall protein content. The phrase “altered fiber content phenotype” refers to measurable phenotype of a genetically modified plant, where the plant displays a statistically significant increase or decrease in overall fiber content (i.e., the percentage of seed mass that is fiber), as compared to the similar, but non-modified plant. A low fiber phenotype refers to decrease in overall fiber 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 or altered 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 or altered plant phenotype or trait, where the modified or altered phenotype or trait is associated with the modified or altered expression of a wild-type polynucleotide sequence or gene.

As used herein, the term “T1” refers to the generation of plants from the seed
5 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.
10 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
15 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
20 cells can be obtained from any plant organ or tissue and cultures prepared therefrom. Provided herein is a transgenic plant cell having an IMQ phenotype and/or an IOQ phenotype. The transgenic plant cell comprises a transformation vector comprising an IMQ nucleotide sequence that encodes or is complementary to a sequence that encodes an IMQ polypeptide. In preferred embodiments, the transgenic plant cell is
25 selected from the group consisting of plants of the *Brassica* species, including canola and rapeseed, soy, corn, sunflower, cotton, cocoa, safflower, oil palm, coconut palm, flax, castor, peanut, wheat, oat and rice. In other embodiments, the plant cell is a seed, pollen, propagule, or embryo cell. The disclosure also provides plant cells from a plant that is the direct progeny or the indirect progeny of a plant
30 grown from said progenitor cells. 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 monocotyledonous and dicotyledonous plants.

As used herein, “transgenic plant” includes a plant that comprises within its genome a heterologous polynucleotide. The heterologous polynucleotide can be
5 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
10 progeny of transformed plants or plant cells that also contain the heterologous polynucleotide are also considered transgenic.

Disclosed herein are transgenic plants having an Improved Seed Quality phenotype. Transgenic plants with an Improved Seed Quality phenotype may include an improved oil quantity and/or an improved meal quality. Transgenic
15 plants with improved meal quality have an IMQ phenotype and transgenic plants with improved oil quantity have an IOQ phenotype. The IMQ phenotype in a transgenic plant may include altered protein and/or fiber content in any part of the transgenic plant, for example in the seeds. The IOQ phenotype in a transgenic plant may include altered oil content in any part of the transgenic plant, for example in the
20 seeds. In particular embodiments, a transgenic plant may include an IOQ phenotype and/or an IMQ phenotype. In some embodiments of a transgenic plant, the IMQ phenotype may be an increase in protein content in the seed and/or a decrease in the fiber content of the seed. In other embodiments of a transgenic plant, the IOQ phenotype is an increase in the oil content of the seed (a high oil phenotype). Also
25 provided is seed meal derived from the seeds of transgenic plants, wherein the seeds have altered protein content and/or altered fiber content. Further provided is oil derived from the seeds of transgenic plants, wherein the seeds have altered oil content. Any of these changes can lead to an increase in the AME from the seed or seed meal from transgenic plants, relative to control, non-transgenic, or wild-type
30 plants. Also provided herein is meal, feed, or food produced from any part of the transgenic plant with an IMQ phenotype and/or IOQ phenotype.

In certain embodiments, the disclosed transgenic plants comprise a transformation vector comprising an IMQ nucleotide sequence that encodes or is complementary to a sequence that encodes an “IMQ” polypeptide. In particular embodiments, expression of an IMQ polypeptide in a transgenic plant causes an altered oil content, an altered protein content, and/or an altered fiber content in the transgenic plant. In preferred embodiments, the transgenic plant is selected from the group consisting of plants of the *Brassica* species, including canola and rapeseed, soy, corn, sunflower, cotton, cocoa, safflower, oil palm, coconut palm, flax, castor, peanut, wheat, oat and rice. Also provided is a method of producing oil or seed meal, comprising growing the transgenic plant and recovering oil and/or seed meal from said plant. The disclosure further provides feed, meal, grain, or seed comprising a nucleic acid sequence that encodes an IMQ polypeptide. The disclosure also provides feed, meal, grain, or seed comprising the IMQ polypeptide, or an ortholog thereof.

Various methods for the introduction of a desired polynucleotide sequence encoding the desired protein into plant cells are available and known to those of skill in the art and include, but are not limited to: (1) physical methods such as microinjection, electroporation, and microprojectile mediated delivery (biolistics or gene gun technology); (2) virus mediated delivery methods; and (3) *Agrobacterium*-mediated transformation methods (see, for example, WO 2007/053482 and WO 2005/107437).

The most commonly used methods for transformation of plant cells are the *Agrobacterium*-mediated DNA transfer process and the biolistics or microprojectile bombardment mediated process (*i.e.*, the gene gun). Typically, nuclear transformation is desired but where it is desirable to specifically transform plastids, such as chloroplasts or amyloplasts, plant plastids may be transformed utilizing a microprojectile-mediated delivery of the desired polynucleotide.

Agrobacterium-mediated transformation is achieved through the use of a genetically engineered soil bacterium belonging to the genus *Agrobacterium*. A number of wild-type and disarmed strains of *Agrobacterium tumefaciens* and *Agrobacterium rhizogenes* harboring Ti or Ri plasmids can be used for gene transfer into plants. Gene transfer is done via the transfer of a specific DNA known as “T-

DNA” that can be genetically engineered to carry any desired piece of DNA into many plant species.

Agrobacterium-mediated genetic transformation of plants involves several steps. The first step, in which the virulent *Agrobacterium* and plant cells are first brought into contact with each other, is generally called “inoculation.” Following the inoculation, the *Agrobacterium* and plant cells/tissues are permitted to be grown together for a period of several hours to several days or more under conditions suitable for growth and T-DNA transfer. This step is termed “co-culture.” Following co-culture and T-DNA delivery, the plant cells are treated with bactericidal or bacteriostatic agents to kill the *Agrobacterium* remaining in contact with the explant and/or in the vessel containing the explant. If this is done in the absence of any selective agents to promote preferential growth of transgenic versus non-transgenic plant cells, then this is typically referred to as the “delay” step. If done in the presence of selective pressure favoring transgenic plant cells, then it is referred to as a “selection” step. When a “delay” is used, it is typically followed by one or more “selection” steps.

With respect to microprojectile bombardment (U.S. Patent No. 5,550,318; U.S. Patent No. 5,538,880, U.S. Patent No. 5,610,042; and PCT Publication WO 95/06128), particles are coated with nucleic acids and delivered into cells by a propelling force. Exemplary particles include those comprised of tungsten, platinum, and preferably, gold.

An illustrative embodiment of a method for delivering DNA into plant cells by acceleration is the Biolistics Particle Delivery System (BioRad, Hercules, CA), which can be used to propel particles coated with DNA or cells through a screen, such as a stainless steel or Nytex screen, onto a filter surface covered with monocot plant cells cultured in suspension.

Microprojectile bombardment techniques are widely applicable, and may be used to transform virtually any plant species. Examples of species that have been transformed by microprojectile bombardment include monocot species such as maize (PCT Publication No. WO 95/06128), barley, wheat (U.S. Patent No. 5,563,055), rice, oat, rye, sugarcane,

and sorghum, as well as a number of dicots including tobacco, soybean (U.S. Patent No. 5,322,783), sunflower, peanut, cotton, tomato, and legumes in general (U.S. Patent No. 5,563,055).

To select or score for transformed plant cells regardless of transformation methodology, the DNA introduced into the cell contains a gene that functions in a regenerative plant tissue to produce a compound that confers upon the plant tissue resistance to an otherwise toxic compound. Genes of interest for use as a selectable, screenable, or scorable marker would include but are not limited to GUS, green fluorescent protein (GFP), luciferase (LUX), antibiotic or herbicide tolerance genes. Examples of antibiotic resistance genes include the penicillins, kanamycin (and neomycin, G418, bleomycin), methotrexate (and trimethoprim), chloramphenicol, and tetracycline. Polynucleotide molecules encoding proteins involved in herbicide tolerance are known in the art, and include, but are not limited to a polynucleotide molecule encoding 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) described in U.S. Patent No. 5,627,061, U.S. Patent No 5,633,435, and U.S. Patent No 6,040,497 and *aroA* described in U.S. Patent No. 5,094,945 for glyphosate tolerance; a polynucleotide molecule encoding bromoxynil nitrilase (Bxn) described in U.S. Patent No. 4,810,648 for Bromoxynil tolerance; a polynucleotide molecule encoding phytoene desaturase (*crtI*) described in Misawa *et al.*, (*Plant J.* 4:833-840, 1993) and Misawa *et al.*, (*Plant J.* 6:481-489, 1994) for norflurazon tolerance; a polynucleotide molecule encoding acetohydroxyacid synthase (AHAS, also known as ALS) described in Sathasiivan *et al.* (*Nucl. Acids Res.* 18:2188-2193, 1990) for tolerance to sulfonylurea herbicides; and the bar gene described in DeBlock, *et al.*, (*EMBO J.* 6:2513-2519, 1987) for glufosinate and bialaphos tolerance.

The regeneration, development, and cultivation of plants from various transformed explants are well documented in the art. This regeneration and growth process typically includes the steps of selecting transformed cells and culturing those individualized cells through the usual stages of embryonic development through the rooted plantlet stage. Transgenic embryos and seeds are similarly regenerated. The resulting transgenic rooted shoots are thereafter planted in an appropriate plant growth medium such as soil. Cells that survive the exposure to the selective agent,

or cells that have been scored positive in a screening assay, may be cultured in media that supports regeneration of plants. Developing plantlets are transferred to soil less plant growth mix, and hardened off, prior to transfer to a greenhouse or growth chamber for maturation.

5 The present invention can be used with any transformable cell or tissue. By transformable as used herein is meant a cell or tissue that is capable of further propagation to give rise to a plant. Those of skill in the art recognize that a number of plant cells or tissues are transformable in which after insertion of exogenous DNA and appropriate culture conditions the plant cells or tissues can form into a
10 differentiated plant. Tissue suitable for these purposes can include but is not limited to immature embryos, scutellar tissue, suspension cell cultures, immature inflorescence, shoot meristem, nodal explants, callus tissue, hypocotyl tissue, cotyledons, roots, and leaves.

Any suitable plant culture medium can be used. Examples of suitable media
15 would include but are not limited to MS-based media (Murashige and Skoog, *Physiol. Plant*, 15:473-497, 1962) or N6-based media (Chu *et al.*, *Scientia Sinica* 18:659, 1975) supplemented with additional plant growth regulators including but not limited to auxins, cytokinins, ABA, and gibberellins. Those of skill in the art are familiar with the variety of tissue culture media, which when supplemented
20 appropriately, support plant tissue growth and development and are suitable for plant transformation and regeneration. These tissue culture media can either be purchased as a commercial preparation, or custom prepared and modified. Those of skill in the art are aware that media and media supplements such as nutrients and growth
25 regulators for use in transformation and regeneration and other culture conditions such as light intensity during incubation, pH, and incubation temperatures that can be optimized for the particular variety of interest.

One of ordinary skill will appreciate that, after an expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard
30 breeding techniques can be used, depending upon the species to be crossed.

Identification of Plants with an Improved Oil Quantity Phenotype and/or Improved Meal Quality Phenotype

An *Arabidopsis* activation tagging screen (ACTTAG) was used to identify the association between 1) ACTTAG plant lines with an altered protein, fiber and/or oil content (phenotype, for example, see columns 4, 5 and 6, respectively, of Table 1, below) and 2) the nucleic acid sequences identified in column 3 of Tables 2 and 3, wherein each nucleic acid sequence is provided with a gene alias or an IMQ designation (IMQ#; see column 1 in Tables 1, 2, and 3). 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, *Plant Physiology*, 122:1003-1013). When the T-DNA inserts into the genome of transformed plants, the enhancer element can cause up-regulation of genes in the vicinity, generally within about nine kilobases (kb) of the enhancers. 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. T1 plants were allowed to grow to maturity, self-fertilize and produce seed. T2 seed was harvested, labeled and stored. To amplify the seed stocks, about eighteen T2 were sown in soil and, after germination, exposed to the selective agent to recover transformed T2 plants. T3 seed from these plants was harvested and pooled. Oil, protein and fiber content of the seed were estimated using Near Infrared Spectroscopy (NIR) as described in the Examples.

Quantitative determination of fatty acid (FA) content (column 7, Table 1) in T2 seeds was performed using the following methods. A sample of 15 to 20 T2 seeds from each line tested. This sample generally contained plants with homozygous insertions, no insertions, and hemizygous insertions in a standard 1:1:2 ratios. The seed sample was massed on UMT-2 ultra-microbalance (Mettler-Toledo Co., Ohio, USA) and then transferred to a glass extraction vial. Lipids were extracted from the seeds and trans-esterified in 500 ul 2.5% H₂SO₄ in MeOH for 3 hours at 80 °C, following the method of Browse *et al.* (Biochem J 235:25-31, 1986) with modifications. A known amount of heptadecanoic acid was included in the reaction as an internal standard. 750 ul of water and 400 ul of hexane were added to

each vial, which was then shaken vigorously and allowed to phase separate. Reaction vials were loaded directly onto gas chromatography (GC) for analysis and the upper hexane phase was sampled by the autosampler. Gas chromatography with Flame Ionization detection was used to separate and quantify the fatty acid methyl esters. Agilent 6890 Plus GC's were used for separation with Agilent Innowax columns (30m x 0.25mm ID, 250um film thickness). The carrier gas was Hydrogen at a constant flow of 2.5 ml/ minute. 1ul of sample was injected in splitless mode (inlet temperature 220°C, Purge flow 15ml/min at 1 minute). The oven was programmed for an initial temperature of 105°C, initial time 0.5 minutes, followed by a ramp of 60°C per minute to 175°C, a 40°C /minute ramp to 260°C with a final hold time of 2 minutes. Detection was by Flame Ionization (Temperature 275°C, Fuel flow 30.0 ml/min, Oxidizer 400.0 ml/min). Instrument control and data collection and analysis were monitored using the Millennium Chromatography Management System (Version 3.2, Waters Corporation, Milford, MA). Peaks were initially identified by comparison with standards. Integration and quantification were performed automatically, but all analyses were subsequently examined manually to verify correct peak identification and acceptable signal to noise ratio before inclusion of the derived results in the study.

The association of an IMQ nucleic acid sequence with an IMQ phenotype or an IOQ phenotype was discovered by analysis of the genomic DNA sequence flanking the T-DNA insertion in the ACTTAG line identified in column 3 of Table 1. An ACTTAG line is a family of plants derived from a single plant that was transformed with a T-DNA element containing four tandem copies of the CaMV 35S enhancers. Accordingly, the disclosed IMQ nucleic acid sequences and/or polypeptides may be employed in the development of transgenic plants having an improved seed quality phenotype, including an IMQ phenotype and/or an IOQ phenotype. IMQ nucleic acid sequences may be used in the generation of transgenic plants, such as oilseed crops, that provide improved oil yield from oilseed processing and result in an increase in the quantity of oil recovered from seeds of the transgenic plant. IMQ nucleic acid sequences may also be used in the generation of transgenic plants, such as feed grain crops, that provide an IMQ phenotype resulting in increased energy for animal feeding, for example, seeds or seed meal with an

altered protein and/or fiber content, resulting in an increase in AME. IMQ nucleic acid sequences may further be used to increase the oil content of specialty oil crops, in order to augment yield and/or recovery of desired unusual fatty acids. Transgenic plants that have been genetically modified to express IMQ polypeptides can be used
 5 in the production of seeds, wherein the transgenic plants are grown, and oil and seed meal are obtained from plant parts (*e.g.* seed) using standard methods.

IMQ Nucleic Acids and Polypeptides

The IMQ designation for each of the IMQ nucleic acid sequences discovered
 10 in the activation tagging screen described herein are listed in column 1 of Tables 1-3, below. The disclosed IMQ polypeptides are listed in column 5 of Table 2 and column 4 of Table 3. As used herein, the term “IMQ polypeptide” refers to any polypeptide that when expressed in a plant causes an IMQ phenotype and/or an IOQ phenotype in any part of the plant, for example the seeds. In one embodiment, an
 15 IMQ polypeptide refers to a full-length IMQ protein, or a fragment, derivative (variant), or ortholog thereof that is “functionally active,” such that the protein fragment, derivative, or ortholog exhibits one or more of the functional activities associated with one or more of the disclosed full-length IMQ polypeptides, for example, the amino acid sequences provided in the GenBank entry referenced in
 20 column 5 of Table 2, which correspond to the amino acid sequences set forth as SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO:
 25 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, SEQ ID NO: 50, SEQ ID NO: 52, SEQ ID NO: 54, SEQ ID NO: 56, SEQ ID NO: 58, SEQ ID NO: 60, SEQ ID NO: 62, SEQ ID NO: 64, SEQ ID NO: 66, SEQ ID NO: 68, SEQ ID NO: 70, SEQ ID NO: 72, SEQ ID NO: 74, SEQ ID NO: 76, SEQ ID NO: 78, SEQ ID NO: 80, SEQ ID NO: 82, SEQ ID NO: 84, SEQ ID NO: 86, SEQ ID
 30 NO: 88, SEQ ID NO: 90, SEQ ID NO: 92, SEQ ID NO: 94, SEQ ID NO: 96, SEQ ID NO: 98, or SEQ ID NO: 100, or an ortholog thereof. In one preferred embodiment, a functionally active IMQ polypeptide causes an IMQ phenotype

and/or an IOQ phenotype in a transgenic plant. In another embodiment, a functionally active IMQ polypeptide causes an altered oil, protein, and/or fiber content phenotype (for example, an altered seed meal content phenotype) when mis-expressed in a plant. In other preferred embodiments, mis-expression of the IMQ polypeptide causes a high oil (such as, increased oil), high protein (such as, increased protein), and/or low fiber (such as, decreased fiber) phenotype in a plant. In another embodiment, mis-expression of the IMQ polypeptide causes an improved AME of meal. In yet another embodiment, a functionally active IMQ polypeptide can rescue defective (including deficient) endogenous IMQ activity when expressed in a plant or in plant cells; the rescuing polypeptide may be from the same or from a different species as the species with the defective polypeptide activity. The disclosure also provides feed, meal, grain, food, or seed comprising the IMQ polypeptide, or a fragment, derivative (variant), or ortholog thereof.

In another embodiment, a functionally active fragment of a full length IMQ polypeptide (for example, a functionally active fragment of a native polypeptide having the amino acid sequence set forth as SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, SEQ ID NO: 50, SEQ ID NO: 52, SEQ ID NO: 54, SEQ ID NO: 56, SEQ ID NO: 58, SEQ ID NO: 60, SEQ ID NO: 62, SEQ ID NO: 64, SEQ ID NO: 66, SEQ ID NO: 68, SEQ ID NO: 70, SEQ ID NO: 72, SEQ ID NO: 74, SEQ ID NO: 76, SEQ ID NO: 78, SEQ ID NO: 80, SEQ ID NO: 82, SEQ ID NO: 84, SEQ ID NO: 86, SEQ ID NO: 88, SEQ ID NO: 90, SEQ ID NO: 92, SEQ ID NO: 94, SEQ ID NO: 96, SEQ ID NO: 98, or SEQ ID NO: 100, or a naturally occurring ortholog thereof) retains one or more of the biological properties associated with the full-length IMQ polypeptide, such as signaling activity, binding activity, catalytic activity, or cellular or extra-cellular localizing activity. An IMQ fragment preferably comprises an IMQ 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 an IMQ protein. Functional domains of IMQ genes are listed in column 8 of Table 2 and can be identified using the PFAM program (Bateman A *et al.*, 1999, *Nucleic Acids Res.* 27:260-262) or INTERPRO (Mulder *et al.*, 2003, *Nucleic Acids Res.* 31, 315-318) program. Functionally active variants of full-length IMQ polypeptides, or fragments thereof, include polypeptides with amino acid insertions, deletions, or substitutions that retain one or more of the biological properties associated with the full-length IMQ polypeptide. In some cases, variants are generated that change the post-translational processing of an IMQ 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 “IMQ nucleic acid” refers to any polynucleotide that when expressed in a plant causes an IMQ phenotype and/or an IOQ phenotype in any part of the plant, for example the seeds. In one embodiment, an IMQ polynucleotide encompasses nucleic acids with the sequence provided in or complementary to the GenBank entry referenced in column 3 of Table 2, which correspond to nucleic acid sequences set forth as SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 23, SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 39, SEQ ID NO: 41, SEQ ID NO: 43, SEQ ID NO: 45, SEQ ID NO: 47, SEQ ID NO: 49, SEQ ID NO: 51, SEQ ID NO: 53, SEQ ID NO: 55, SEQ ID NO: 57, SEQ ID NO: 59, SEQ ID NO: 61, SEQ ID NO: 63, SEQ ID NO: 65, SEQ ID NO: 67, SEQ ID NO: 69, SEQ ID NO: 71, SEQ ID NO: 73, SEQ ID NO: 75, SEQ ID NO: 77, SEQ ID NO: 79, SEQ ID NO: 81, SEQ ID NO: 83, SEQ ID NO: 85, SEQ ID NO: 87, SEQ ID NO: 89, SEQ ID NO: 91, SEQ ID NO: 93, SEQ ID NO: 95, SEQ ID NO: 97, or SEQ ID NO: 99, as well as functionally active fragments, derivatives, or orthologs thereof. An IMQ nucleic acid of this disclosure may be DNA, derived from genomic DNA or cDNA, or RNA. Genomic sequences of the genes listed in Table 2 are known and available in public databases such as GenBank.

In one embodiment, a functionally active IMQ nucleic acid encodes or is complementary to a nucleic acid that encodes a functionally active IMQ

polypeptide. A functionally active IMQ nucleic acid also includes 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 IMQ polypeptide. An IMQ 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 IMQ polypeptide, or an intermediate form. An IMQ 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 IMQ nucleic acid is capable of being used in the generation of loss-of-function IMQ phenotypes, for instance, via antisense suppression, co-suppression, etc. The disclosure also provides feed, meal, grain, food, or seed comprising a nucleic acid sequence that encodes an IMQ polypeptide.

In one preferred embodiment, an IMQ nucleic acid used in the disclosed methods comprises a nucleic acid sequence that encodes, or is complementary to a sequence that encodes, an IMQ polypeptide having at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, or 99% sequence identity to a disclosed IMQ polypeptide sequence, for example the amino acid sequence set forth as SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, SEQ ID NO: 50, SEQ ID NO: 52, SEQ ID NO: 54, SEQ ID NO: 56, SEQ ID NO: 58, SEQ ID NO: 60, SEQ ID NO: 62, SEQ ID NO: 64, SEQ ID NO: 66, SEQ ID NO: 68, SEQ ID NO: 70, SEQ ID NO: 72, SEQ ID NO: 74, SEQ ID NO: 76, SEQ ID NO: 78, SEQ ID NO: 80, SEQ ID NO: 82, SEQ ID NO: 84, SEQ ID NO: 86, SEQ ID

NO: 88, SEQ ID NO: 90, SEQ ID NO: 92, SEQ ID NO: 94, SEQ ID NO: 96, SEQ ID NO: 98, or SEQ ID NO: 100.

In another embodiment, an IMQ polypeptide comprises a polypeptide sequence with at least 50% or 60% identity to a disclosed IMQ polypeptide sequence (for example, the amino acid sequence set forth as SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, SEQ ID NO: 50, SEQ ID NO: 52, SEQ ID NO: 54, SEQ ID NO: 56, SEQ ID NO: 58, SEQ ID NO: 60, SEQ ID NO: 62, SEQ ID NO: 64, SEQ ID NO: 66, SEQ ID NO: 68, SEQ ID NO: 70, SEQ ID NO: 72, SEQ ID NO: 74, SEQ ID NO: 76, SEQ ID NO: 78, SEQ ID NO: 80, SEQ ID NO: 82, SEQ ID NO: 84, SEQ ID NO: 86, SEQ ID NO: 88, SEQ ID NO: 90, SEQ ID NO: 92, SEQ ID NO: 94, SEQ ID NO: 96, SEQ ID NO: 98, or SEQ ID NO: 100) and may have at least 70%, 80%, 85%, 90%, 95%, 97%, 98%, or 99% sequence identity to a disclosed IMQ polypeptide sequence. In a further embodiment, an IMQ polypeptide comprises 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 98%, or 99% sequence identity to a disclosed IMQ polypeptide sequence, and may include a conserved protein domain of the IMQ polypeptide (such as the protein domain(s) listed in column 8 of Table 2). In another embodiment, an IMQ polypeptide comprises a polypeptide sequence with at least 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 98%, or 99% sequence identity to a functionally active fragment of the polypeptide referenced in column 5 of Table 2. In yet another embodiment, an IMQ polypeptide comprises a polypeptide sequence with at least 50%, 60 %, 70%, 80%, 90%, 95%, 97%, 98%, or 99% identity to the polypeptide sequence of the GenBank entry referenced in column 5 of Table 2 over its entire length and comprises a conserved protein domain(s) listed in column 8 of Table 2.

In another aspect, an IMQ polynucleotide sequence is at least 50% to 60% identical over its entire length to a disclosed IMQ nucleic acid sequence, such as the nucleic acid sequence set forth as SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15,

SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 23, SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 39, SEQ ID NO: 41, SEQ ID NO: 43, SEQ ID NO: 45, SEQ ID NO: 47, SEQ ID NO: 49, SEQ ID NO: 51, SEQ ID NO: 53, SEQ ID NO: 55, SEQ ID NO: 57, SEQ ID NO: 59, SEQ ID NO: 61, SEQ ID NO: 63, SEQ ID NO: 65, SEQ ID NO: 67, SEQ ID NO: 69, SEQ ID NO: 71, SEQ ID NO: 73, SEQ ID NO: 75, SEQ ID NO: 77, SEQ ID NO: 79, SEQ ID NO: 81, SEQ ID NO: 83, SEQ ID NO: 85, SEQ ID NO: 87, SEQ ID NO: 89, SEQ ID NO: 91, SEQ ID NO: 93, SEQ ID NO: 95, SEQ ID NO: 97, or SEQ ID NO: 99, or nucleic acid sequences that are complementary to such an IMQ sequence, and may comprise at least 70%, 80%, 85%, 90%, 95%, 97%, 98%, or 99% sequence identity to the disclosed IMQ sequence, or a functionally active fragment thereof, or complementary sequences. In another embodiment, a disclosed IMQ nucleic acid comprises a nucleic acid sequence as shown in SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 23, SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 39, SEQ ID NO: 41, SEQ ID NO: 43, SEQ ID NO: 45, SEQ ID NO: 47, SEQ ID NO: 49, SEQ ID NO: 51, SEQ ID NO: 53, SEQ ID NO: 55, SEQ ID NO: 57, SEQ ID NO: 59, SEQ ID NO: 61, SEQ ID NO: 63, SEQ ID NO: 65, SEQ ID NO: 67, SEQ ID NO: 69, SEQ ID NO: 71, SEQ ID NO: 73, SEQ ID NO: 75, SEQ ID NO: 77, SEQ ID NO: 79, SEQ ID NO: 81, SEQ ID NO: 83, SEQ ID NO: 85, SEQ ID NO: 87, SEQ ID NO: 89, SEQ ID NO: 91, SEQ ID NO: 93, SEQ ID NO: 95, SEQ ID NO: 97, or SEQ ID NO: 99,, or nucleic acid sequences that are complementary to such an IMQ sequence, and nucleic acid sequences that have substantial sequence homology to a such IMQ sequences. As used herein, the phrase “substantial sequence homology” refers to those nucleic acid sequences that have slight or inconsequential sequence variations from such IMQ sequences, *i.e.*, the sequences function in substantially the same manner and encode an IMQ polypeptide.

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 an identified 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.*,
5 *J. Mol. Biol.*, 1990, 215:403-410) 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 “percent (%) identity value” is determined by the number of
10 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 performing 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
15 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
20 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
25 include sequences that selectively hybridize to the disclosed IMQ nucleic acid sequences (for example, the nucleic acid sequence set forth as SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 23, SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31,
30 SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 39, SEQ ID NO: 41, SEQ ID NO: 43, SEQ ID NO: 45, SEQ ID NO: 47, SEQ ID NO: 49, SEQ ID NO: 51, SEQ ID NO: 53, SEQ ID NO: 55, SEQ ID NO: 57, SEQ ID NO: 59, SEQ

ID NO: 61, SEQ ID NO: 63, SEQ ID NO: 65, SEQ ID NO: 67, SEQ ID NO: 69,
 SEQ ID NO: 71, SEQ ID NO: 73 SEQ ID NO: 75, SEQ ID NO: 77, SEQ ID NO:
 79, SEQ ID NO: 81, SEQ ID NO: 83, SEQ ID NO: 85, SEQ ID NO: 87, SEQ ID
 NO: 89, SEQ ID NO: 91, SEQ ID NO: 93, SEQ ID NO: 95, SEQ ID NO: 97, or
 5 SEQ ID NO: 99). 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.*, 1989, *Molecular Cloning: A Laboratory Manual*
 10 (Second Edition), Cold Spring Harbor Press, Plainview, N.Y.).

In some embodiments, a nucleic acid molecule of the disclosure is capable of
 hybridizing to a nucleic acid molecule containing the disclosed nucleotide sequence
 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
 15 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
 20 (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%
 25 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%
 30 formamide, 5 x SSC, 50 mM sodium phosphate (pH 7.6), 5X Denhardt's solution,
 10% dextran sulfate, and 20 µg/ml denatured sheared salmon sperm DNA;

hybridization in the same buffer for 18 to 20 hours; and washing of filters in 1 x SSC at about 37° C for 1 hour.

As a result of the degeneracy of the genetic code, a number of polynucleotide sequences encoding an IMQ polypeptide can be produced. For example, codons may
5 be selected to increase the rate at which expression of the polypeptide occurs in a particular host species, in accordance with the optimum codon usage dictated by the particular host organism (see, e.g., Nakamura *et al.*, 1999, *Nucleic Acids Res.* 27:292). Such sequence variants may be used in the methods disclosed herein.

The disclosed methods may use orthologs of a disclosed *Arabidopsis* IMQ
10 nucleic acid sequence. Representative putative orthologs of each of the disclosed *Arabidopsis* IMQ genes are identified in column 3 of Table 3, below. Methods of identifying the orthologs in other plant species are known in the art. In general, 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
15 duplication event follows speciation, a single gene in one species, such as *Arabidopsis*, may correspond to multiple genes (paralogs) in another. As used herein, the term “orthologs” encompasses paralogs. When sequence data is available for a particular plant species, orthologs are generally identified by sequence homology analysis, such as BLAST analysis, usually using protein bait
20 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, 1998, *Proc. Natl. Acad. Sci.*, 95:5849-5856; Huynen MA *et al.*, 2000, *Genome Research*, 10:1204-1210).

Programs for multiple sequence alignment, such as CLUSTAL (Thompson
25 JD *et al.*, 1994, *Nucleic Acids Res.* 22:4673-4680) may be used to highlight conserved regions and/or residues of orthologous proteins and to generate phylogenetic trees. In a phylogenetic tree representing multiple homologous sequences from diverse species (e.g., retrieved through BLAST analysis), orthologous sequences from two species generally appear closest on the tree with
30 respect to all other sequences from these two species. Structural threading or other analysis of protein folding (e.g., using software by ProCeryon, Biosciences, Salzburg, Austria) may also identify potential orthologs. Nucleic acid hybridization

methods may also be used to find orthologous genes and are preferred when sequence data are not available. Degenerate PCR and screening of cDNA or genomic DNA libraries are common methods for finding related gene sequences and are well known in the art (see, *e.g.*, Sambrook, 1989, *Molecular Cloning: A Laboratory Manual* (Second Edition), Cold Spring Harbor Press, Plainview, N.Y.; Dieffenbach and Dveksler, 1995, *PCR Primer: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, NY). For instance, methods for generating a cDNA library from the plant species of interest and probing the library with partially homologous gene probes are described in Sambrook *et al.* A highly conserved portion of the *Arabidopsis* IMQ coding sequence may be used as a probe. IMQ ortholog nucleic acids may hybridize to the nucleic acid of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 23, SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 39, SEQ ID NO: 41, SEQ ID NO: 43, SEQ ID NO: 45, SEQ ID NO: 47, SEQ ID NO: 49, SEQ ID NO: 51, SEQ ID NO: 53, SEQ ID NO: 55, SEQ ID NO: 57, SEQ ID NO: 59, SEQ ID NO: 61, SEQ ID NO: 63, SEQ ID NO: 65, SEQ ID NO: 67, SEQ ID NO: 69, SEQ ID NO: 71, SEQ ID NO: 73, SEQ ID NO: 75, SEQ ID NO: 77, SEQ ID NO: 79, SEQ ID NO: 81, SEQ ID NO: 83, SEQ ID NO: 85, SEQ ID NO: 87, SEQ ID NO: 89, SEQ ID NO: 91, SEQ ID NO: 93, SEQ ID NO: 95, SEQ ID NO: 97, or SEQ ID NO: 99 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 DNA clone.

Alternatively, it is possible to initiate an EST project to generate a database of sequence information for the plant species of interest. In another approach, antibodies that specifically bind known IMQ polypeptides are used for ortholog isolation (see, *e.g.*, Harlow and Lane, 1988, 1999, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York). Western blot analysis can determine that an IMQ ortholog (*i.e.*, a protein orthologous to a disclosed IMQ polypeptide) 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
5 candidate ortholog(s) are identified by any of these means, candidate orthologous sequence are used as bait (the “query”) for the reverse BLAST against sequences from *Arabidopsis* or other species in which IMQ nucleic acid and/or polypeptide sequences have been identified.

IMQ nucleic acids and polypeptides may be obtained using any available
10 method. For instance, techniques for isolating cDNA or genomic DNA sequences of interest by screening DNA libraries or by using polymerase chain reaction (PCR), as previously described, are well known in the art. Alternatively, nucleic acid sequence may be synthesized. Any known method, such as site directed mutagenesis (Kunkel
TA *et al.*, 1991, *Methods Enzymol.* 204:125-39), may be used to introduce desired
15 changes into a cloned nucleic acid.

In general, the methods disclosed herein involve incorporating the desired form of the IMQ nucleic acid into a plant expression vector for transformation of plant cells, and the IMQ polypeptide is expressed in the host plant. Transformed plants and plant cells expressing an IMQ polypeptide express an IMQ phenotype
20 and/or an IOQ phenotype and, in one specific, non-limiting example, may have high (increased) oil, high (increased) protein, and/or low (decreased) fiber content.

An “isolated” IMQ 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
25 natural source of the IMQ nucleic acid. However, an isolated IMQ nucleic acid molecule includes IMQ nucleic acid molecules contained in cells that ordinarily express IMQ where, for example, the nucleic acid molecule is in a chromosomal location different from that of natural cells.

Generation of Genetically Modified Plants with an Improved Oil Quantity**Phenotype and/or an Improved Meal Quality Phenotype**

The disclosed IMQ nucleic acids and polypeptides may be used in the generation of transgenic plants having a modified or altered oil, protein, and/or fiber content phenotype. As used herein, an “altered oil content (phenotype)” may refer to altered oil content in any part of the plant. In a preferred embodiment, altered expression of the IMQ gene in a plant is used to generate plants with a high oil content (phenotype). As used herein, an “altered protein content (phenotype)” may refer to altered protein content in any part of the plant. In a preferred embodiment, altered expression of the IMQ gene in a plant is used to generate plants with a high (or increased) protein content (phenotype). As used herein, an “altered fiber content (phenotype)” may refer to altered fiber content in any part of the plant. In a preferred embodiment, altered expression of the IMQ gene in a plant is used to generate plants with a low (or decreased) fiber content (phenotype). The altered oil, protein, and/or fiber content is often observed in seeds. Examples of a transgenic plant include plants comprising a plant transformation vector with a nucleotide sequence that encodes or is complementary to a sequence that encodes an IMQ polypeptide having the amino acid sequence as set forth in SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, SEQ ID NO: 50, SEQ ID NO: 52, SEQ ID NO: 54, SEQ ID NO: 56, SEQ ID NO: 58, SEQ ID NO: 60, SEQ ID NO: 62, SEQ ID NO: 64, SEQ ID NO: 66, SEQ ID NO: 68, SEQ ID NO: 70, SEQ ID NO: 72, SEQ ID NO: 74, SEQ ID NO: 76, SEQ ID NO: 78, SEQ ID NO: 80, SEQ ID NO: 82, SEQ ID NO: 84, SEQ ID NO: 86, SEQ ID NO: 88, SEQ ID NO: 90, SEQ ID NO: 92, SEQ ID NO: 94, SEQ ID NO: 96, SEQ ID NO: 98, or SEQ ID NO: 100, or an ortholog thereof.

Transgenic plants, such as corn, soybean and canola containing the disclosed nucleic acid sequences, can be used in the production of vegetable oil and meal. Vegetable oil is used in a variety of food products, while meal from seed is used as

an animal feed. After harvesting seed from transgenic plants, the seed is cleaned to remove plant stalks and other material and then flaked in roller mills to break the hulls. The crushed seed is heated to 75-100° C to denature hydrolytic enzymes, lyse the unbroken oil containing cells, and allow small oil droplets to coalesce. Most of the oil is then removed (and can be recovered) by pressing the seed material in a screw press. The remaining oil is removed from the presscake by extraction with and organic solvents, such as hexane. The solvent is removed from the meal by heating it to approximately 100° C. After drying, the meal is then granulated to a consistent form. The meal, containing the protein, digestible carbohydrate, and fiber of the seed, may be mixed with other materials prior to being used as an animal feed.

The methods described herein for generating transgenic plants are generally applicable to all plants. Although activation tagging and gene identification is carried out in *Arabidopsis*, the IMQ nucleic acid sequence (or an ortholog, variant or fragment thereof) may be expressed in any type of plant. In a preferred embodiment, oil-producing plants 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 (*Theobroma cacao*), safflower (*Carthamus tinctorius*), oil palm (*Elaeis guineensis*), coconut palm (*Cocos nucifera*), flax (*Linum usitatissimum*), castor (*Ricinus communis*), and peanut (*Arachis hypogaea*), as well as wheat, rice and oat. Fruit- and vegetable-bearing plants, grain-producing plants, nut-producing plants, rapid cycling *Brassica* species, alfalfa (*Medicago sativa*), tobacco (*Nicotiana*), turfgrass (*Poaceae* family), other forage crops, and wild species may also be a source of unique fatty acids. In other embodiments, any plant expressing the IMQ nucleic acid sequence can also express increased protein and/or decreased fiber content in a specific plant part or organ, such as in seeds.

The skilled artisan will recognize that a wide variety of transformation techniques exist in the art, and new techniques are continually becoming available. Any technique that is suitable for the target host plant can be employed within the scope of the present invention. For example, the constructs can be introduced in a variety of forms including, but not limited to, as a strand of DNA, in a plasmid, or in

an artificial chromosome. The introduction of the constructs into the target plant cells can be accomplished by a variety of techniques, including, but not limited to, *Agrobacterium*-mediated transformation, electroporation, microinjection, microprojectile bombardment, calcium-phosphate-DNA co-precipitation, or liposome-mediated transformation of a heterologous nucleic acid. The transformation of the plant is preferably permanent, *i.e.* by integration of the introduced expression constructs into the host plant genome, so that the introduced constructs are passed onto successive plant generations. Depending upon the intended use, a heterologous nucleic acid construct comprising an IMQ polynucleotide may encode the entire protein or a biologically active portion thereof.

In one embodiment, binary Ti-based vector systems may be used to transfer polynucleotides. Standard *Agrobacterium* binary vectors are known to those of skill in the art, and many are commercially available (*e.g.*, pBI121 Clontech Laboratories, Palo Alto, CA). A construct or vector may include a plant promoter to express the nucleic acid molecule of choice. In a preferred embodiment, the promoter is a plant promoter.

The optimal procedure for transformation of plants with *Agrobacterium* vectors will vary with the type of plant being transformed. Exemplary methods for *Agrobacterium*-mediated transformation include transformation of explants of hypocotyl, shoot tip, stem or leaf tissue, derived from sterile seedlings and/or plantlets. Such transformed plants may be reproduced sexually, or by cell or tissue culture. *Agrobacterium* transformation has been previously described for a large number of different types of plants and methods for such transformation may be found in the scientific literature. Of particular relevance are methods to transform commercially important crops, such as plants of the *Brassica* species, including canola and rapeseed, (De Block *et al.*, 1989, *Plant Physiol.*, 91:694-701), sunflower (Everett *et al.*, 1987, *Bio/Technology*, 5:1201), soybean (Christou *et al.*, 1989, *Proc. Natl. Acad. Sci USA*, 86:7500-7504; Kline *et al.*, 1987, *Nature*, 327:70), wheat, rice and oat.

Expression (including transcription and translation) of an IMQ nucleic acid sequence 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 an IMQ 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

5 promoters include the raspberry E4 promoter (U.S. Patent Nos. 5,783,393 and 5,783,394), the nopaline synthase (NOS) promoter (Ebert *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 84:5745-5749, 1987), the octopine synthase (OCS) promoter (which is carried on tumor-inducing plasmids of *Agrobacterium tumefaciens*), the caulimovirus promoters such as the cauliflower mosaic virus (CaMV) 19S promoter

10 (Lawton *et al.*, *Plant Mol. Biol.* 9:315-324, 1987) and the CaMV 35S promoter (Odell *et al.*, *Nature* 313:810-812, 1985 and Jones JD *et al.*, 1992, *Transgenic Res.*, 1:285-297), the figwort mosaic virus 35S-promoter (U.S. Patent No. 5,378,619), the light-inducible promoter from the small subunit of ribulose-1,5-bis-phosphate carboxylase (ssRUBISCO), the Adh promoter (Walker *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 84:6624-6628, 1987), the sucrose synthase promoter (Yang *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 87:4144-4148, 1990), the R gene complex promoter (Chandler *et al.*, *The Plant Cell* 1:1175-1183, 1989), the chlorophyll a/b binding protein gene promoter, the CsVMV promoter (Verdaguer B *et al.*, 1998, *Plant Mol Biol.*, 37:1055-1067), and the melon actin promoter (published PCT application

15 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, *Plant Mol Bio.*, 21:625-640).

In one preferred embodiment, expression of the IMQ nucleic acid sequence is under control of regulatory sequences from genes whose expression is associated

25 with early seed and/or embryo development. Indeed, in a preferred embodiment, the promoter used is a seed-enhanced promoter. Examples of such promoters include the 5' regulatory regions from such genes as napin (Kridl *et al.*, *Seed Sci. Res.* 1:209:219, 1991), globulin (Belanger and Kriz, *Genet.*, 129: 863-872, 1991, GenBank Accession No. L22295), gamma zein Z 27 (Lopes *et al.*, *Mol Gen Genet.*, 247:603-613, 1995), L3 oleosin promoter (U.S. Patent No. 6,433,252), phaseolin (Bustos *et al.*, *Plant Cell*, 1(9):839-853, 1989), arcelin5 (U.S. Application No. 2003/0046727), a soybean 7S promoter, a 7S α promoter (U.S. Application No.

30

2003/0093828), the soybean 7S α ' beta conglycinin promoter, a 7S α ' promoter (Beachy *et al.*, *EMBO J.*, 4:3047, 1985; Schuler *et al.*, *Nucleic Acid Res.*, 10(24):8225-8244, 1982), soybean trypsin inhibitor (Riggs *et al.*, *Plant Cell* 1(6):609-621, 1989), ACP (Baerson *et al.*, *Plant Mol. Biol.*, 22(2):255-267, 1993),
5 stearyl-ACP desaturase (Slocombe *et al.*, *Plant Physiol.* 104(4):167-176, 1994), soybean α ' subunit of β -conglycinin (Chen *et al.*, *Proc. Natl. Acad. Sci.* 83:8560-8564, 1986), *Vicia faba* USP (P-Vf.Usp, SEQ ID NO: 1, 2, and 3 in (U.S. Application No. 2003/229918) and *Zea mays* L3 oleosin promoter (Hong *et al.*, *Plant Mol. Biol.*, 34(3):549-555, 1997). Also included are the zeins, which are a
10 group of storage proteins found in corn endosperm. Genomic clones for zein genes have been isolated (Pedersen *et al.*, *Cell*, 29:1015-1026, 1982; and Russell *et al.*, *Transgenic Res.* 6(2):157-168) and the promoters from these clones, including the 15 kD, 16 kD, 19 kD, 22 kD, 27 kD and genes, could also be used. Other promoters known to function, for example, in corn include the promoters for the following
15 genes: *waxy*, *Brittle*, *Shrunken 2*, Branching enzymes I and II, starch synthases, debranching enzymes, oleosins, glutelins and sucrose synthases. 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-
20 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).
25 Cereal genes whose promoters are associated with early seed and embryo development include rice *glutelin* ("GluA-3," Yoshihara and Takaiwa, 1996, *Plant Cell Physiol.* 37:107-11; "GluB-1," Takaiwa *et al.*, 1996, *Plant Mol. Biol.* 30:1207-21; Washida *et al.*, 1999, *Plant Mol. Biol.* 40:1-12; "Gt3," Leisy *et al.*, 1990, *Plant Mol. Biol.* 14:41-50), rice *prolamin* (Zhou & Fan, 1993, *Transgenic Res.* 2:141-6),
30 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.*, U.S. Patent 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). Additional promoters that may be utilized are described, for example, in U.S. Patent Nos. 5,378,619; 5,391,725; 5,428,147; 5,447,858; 5,608,144; 5,608,144; 5,614,399; 5,633,441; 5,633,435; and 4,633,436.

In yet another aspect, in some cases it may be desirable to inhibit the expression of the endogenous IMQ nucleic acid sequence 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, *Nature*, 334:724-726; van der Krol *et al.*, 1988, *BioTechniques*, 6:958-976); co-suppression (Napoli, *et al.*, 1990, *Plant Cell*, 2:279-289); ribozymes (PCT Publication WO 97/10328); and combinations of sense and antisense (Waterhouse, *et al.*, 1998, *Proc. Natl. Acad. Sci. USA*, 95:13959-13964). 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, *Proc. Natl. Acad. Sci. USA*, 85:8805-8809), a partial cDNA sequence including fragments of 5' coding sequence, (Cannon *et al.*, 1990, *Plant Mol. Biol.*, 15:39-47), or 3' non-coding sequences (Ch'ng *et al.*, 1989, *Proc. Natl. Acad. Sci. USA*, 86:10006-10010).

Cosuppression techniques may use the entire cDNA sequence (Napoli *et al.*, 1990, *Plant Cell*, 2:279-289; van der Krol *et al.*, 1990, *Plant Cell*, 2:291-299), or a partial cDNA sequence (Smith *et al.*, 1990, *Mol. Gen. Genetics*, 224:477-481).

Standard molecular and genetic tests may be performed to further analyze the association between a nucleic acid sequence 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 over-expression, 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, *Arch. Virol. Suppl.* 15:189-201).

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, *Cur. Opin. Plant Biol.* 2(2):96-103; 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 over-expression 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.

5

Generation of Mutated Plants with an Improved Oil Quantity Phenotype and/or Improved Meal Quality Phenotype

Additional methods are disclosed herein of generating a plant having an IMQ and/or an IOQ phenotype, wherein a plant is identified that has an allele in its IMQ
10 nucleic acid sequence that results in an IMQ phenotype and/or an IOQ phenotype, compared to plants lacking the allele. The plant can generate progeny, wherein the progeny inherit the allele and have an IMQ phenotype and/or an IOQ phenotype. For example, provided herein is a method of identifying plants that have mutations in the endogenous IMQ nucleic acid sequence that confer an IMQ phenotype and/or
15 an IOQ phenotype and generating progeny of these plants with an IMQ and/or IOQ phenotype that are not genetically modified. In some embodiments, the plants have an IMQ phenotype with an altered protein and/or fiber content or seed meal content, or an IOQ phenotype, with an altered oil content.

In one method, called “TILLING” (for targeting induced local lesions in
20 genomes), mutations are induced in the seed of a plant of interest, for example, using EMS (ethylmethane sulfonate) treatment. The resulting plants are grown and self-fertilized, and the progeny are used to prepare DNA samples. PCR amplification and sequencing of the IMQ nucleic acid sequence is used to identify whether a mutated plant has a mutation in the IMQ nucleic acid sequence. Plants having IMQ
25 mutations may then be tested for altered oil, protein, and/or fiber content, or alternatively, plants may be tested for altered oil, protein, and/or fiber content, and then PCR amplification and sequencing of the IMQ nucleic acid sequence is used to determine whether a plant having altered oil, protein, and/or fiber content has a mutated IMQ nucleic acid sequence. TILLING can identify mutations that may alter
30 the expression of specific genes or the activity of proteins encoded by these genes (see Colbert *et al.*, 2001, *Plant Physiol.* 126:480-484; McCallum *et al.*, 2000, *Nature Biotechnology* 18:455-457).

In another method, a candidate gene/Quantitative Trait Locus (QTLs) approach can be used in a marker-assisted breeding program to identify alleles of or mutations in the IMQ nucleic acid sequence or orthologs of the IMQ nucleic acid sequence that may confer altered oil, protein, and/or fiber content (see Bert *et al.*,
 5 *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 disclosure, an IMQ nucleic acid is used to identify whether a plant having altered oil, protein, and/or fiber content has a mutation an endogenous IMQ nucleic acid sequence or has a particular allele that causes altered oil, protein, and/or fiber content.

10 The scope of the claims should not be limited by the preferred embodiments set forth in the examples, but should be given the broadest interpretation consistent with the description as a whole.

15 EXAMPLES

EXAMPLE 1

Generation of Plants with an IMQ Phenotype and/or an IOQ Phenotype by Transformation with an Activation Tagging Construct

20 This Example describes the generation of transgenic plants with altered oil, protein, and/or fiber content.

Mutants were generated using the activation tagging “ACTTAG” vector, pSKI015 (GI#6537289; Weigel D *et al.*, 2000, *Plant Physiology*, 122:1003-1013). Standard methods were used for the generation of *Arabidopsis* transgenic plants, and
 25 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
 30 herbicide resistance. T2 seed (from T1 plants) was harvested and sown in soil. T2

plants were exposed to the herbicide to kill plants lacking the ACTTAG vector. T2 plants were grown to maturity, allowed to self-fertilize and set seed. T3 seed (from the T2 plants) was harvested in bulk for each line.

T3 seed was analyzed by Near Infrared Spectroscopy (NIR) at the time of
5 harvest. NIR spectra were captured using a Bruker 22 near infrared spectrometer. Bruker Software was used to estimate total seed oil, total seed protein and total seed fiber content using data from NIR analysis and reference methods according to the manufacturer's instructions. Oil content predicting calibrations were developed following the general method of AOCS Procedure Am1-92, Official Methods and
10 Recommended Practices of the American Oil Chemists Society, 5th Ed., AOCS, Champaign, Ill. A NIR protein content predicting calibration was developed using total nitrogen content data of seed samples following the general method of Dumas Procedure AOAC 968.06 (Official Methods of Analysis of AOAC International 17th Edition AOAC, Gaithersburg, MD). A fiber content predicting calibration was
15 developed by measuring crude fiber content in a set of seed samples. Fiber content of in a known mass of seed was determined using the method of Honig and Rackis, (1979, *J. Agri. Food Chem.*, 27: 1262-1266). Digestible protein content of in a known mass of seed was determined by quantifying the individual amino acids liberated by an acid hydrolysis Steine and Moore (1958, *Anal. Chem.*, 30:1185-
20 1190). The quantification was performed by the Amino Quant (Agilent). The undigested protein remaining associated with the non digestible fraction is measured by the same method described for the whole seed homogenate. Digestible protein content is determined by subtracting the amount of undigested protein associated with the non digestible fraction from the total amount of protein in the seed sample.

25 Seed oil, protein, digestible protein and fiber values in 82,274 lines were determined by NIR spectroscopy and normalized to allow comparison of seed component values in plants grown at different times. Oil, protein and fiber values were normalized by calculating the average oil, protein and fiber values in seed from all plants planted on the same day (including a large number of other ACTTAG
30 plants, including control, wild-type, or non-transgenic plants). The seed components for each line was expressed as a "percent relative value" which was calculated by dividing the component value for each line with the average component value for all

lines planted on the same day (which should approximate the value in control, wild-type, or non-transgenic plants). The “percent relative protein” and “percent relative fiber” were calculated similarly.

Inverse PCR was used to recover genomic DNA flanking the T-DNA insertion. The PCR product was subjected to sequence analysis and placed on the genome using a basic BLASTN search and/or a search of the *Arabidopsis* Information Resource (TAIR) database (available at the publicly available website). Promoters within 9 kb of the enhancers in the ACTTAG element are considered to be within “activation space.” Genes with T-DNA inserts within coding sequences were not considered to be within “activation space.” The ACTTAG lines with the above average oil and protein values, and below average fiber values were identified and are listed in column 3 of Table 1.

Table 1

1. Gene alias	2. Tair	3. ACTTAG Line	4. Relative Seed Protein Content	5. Relative Seed Fiber Content	6. Relative Seed Oil Content	7. GC FA
IMQ21.1	At1g05230	W000032537	114.20%	100.35%	79.63%	
IMQ21.1	At1g05230	W000032537	114.20%	100.35%	79.63%	
IMQ21.2	At1g05240	W000032537	114.20%	100.35%	79.63%	
IMQ21.3	At1g05250	W000032537	114.20%	100.35%	79.63%	
IMQ21.4	At1g05260	W000032537	114.20%	100.35%	79.63%	
IMQ22.1	At1g10140	W000181281	106.75%	91.28%	94.79%	
IMQ22.2	At1g10150	W000181281	106.75%	91.28%	94.79%	
IMQ22.3	At1g10155	W000181281	106.75%	91.28%	94.79%	
IMQ23.1	At1g13630	W000189711	156.76%	98.48%	62.78%	
IMQ23.2	At1g13640	W000189711	156.76%	98.48%	62.78%	
IMQ23.3	At1g13650	W000189711	156.76%	98.48%	62.78%	
IMQ23.3	At1g13650	W000189711	156.76%	98.48%	62.78%	
IMQ24.1	At1g25400	W000092974	113.43%	93.89%	92.75%	
IMQ24.2	At1g25410	W000092974	113.43%	93.89%	92.75%	
IMQ24.3	At1g25420	W000092974	113.43%	93.89%	92.75%	
IMQ24.3	At1g25420	W000092974	113.43%	93.89%	92.75%	
IMQ24.3	At1g25420	W000092974	113.43%	93.89%	92.75%	
IMQ25.1	At1g27630	W000156857	115.45%	91.17%	93.30%	
IMQ25.2	At1g27640	W000156857	115.45%	91.17%	93.30%	
IMQ25.3	At1g27650	W000156857	115.45%	91.17%	93.30%	
IMQ25.4	At1g27660	W000156857	115.45%	91.17%	93.30%	94.87%
IMQ26.1	At1g34160	W000181882	106.61%	89.49%	95.13%	
IMQ26.2	At1g34170	W000181882	106.61%	89.49%	95.13%	
IMQ26.2	At1g34170	W000181882	106.61%	89.49%	95.13%	
IMQ26.3	At1g34180	W000181882	106.61%	89.49%	95.13%	

1. Gene alias	2. Tair	3. ACTTAG Line	4. Relative Seed Protein Content	5. Relative Seed Fiber Content	6. Relative Seed Oil Content	7. GC FA
IMQ27.1	At1g45160	W000144884	99.69%	99.35%	106.43%	
IMQ27.2	At1g45170	W000144884	99.69%	99.35%	106.43%	
IMQ27.3	At1g45180	W000144884	99.69%	99.35%	106.43%	
IMQ28.1	At1g52140	W000160181	111.10%	92.26%	96.12%	
IMQ28.2	At1g52150	W000160181	111.10%	92.26%	96.12%	
IMQ28.2	At1g52150	W000160181	111.10%	92.26%	96.12%	
IMQ29.1	At1g58410	W000050532	100.73%	91.01%	101.59%	108.07%
IMQ29.1	At1g58410	W000050532	100.73%	91.01%	101.59%	
IMQ29.1	At1g58410	W000050532	100.73%	91.01%	101.59%	
IMQ29.2	At1g58420	W000050532	100.73%	91.01%	101.59%	
IMQ29.3	At1g58430	W000050532	100.73%	91.01%	101.59%	
IMQ30.1	At1g75670	W000161468	114.49%	91.60%	93.46%	
IMQ30.1	At1g75670	W000161468	114.49%	91.60%	93.46%	
IMQ30.2	At1g75680	W000161468	114.49%	91.60%	93.46%	
IMQ30.3	At1g75690	W000161468	114.49%	91.60%	93.46%	
IMQ30.4	At1g75700	W000161468	114.49%	91.60%	93.46%	
IMQ31.1	At1g77730	W000032887	115.30%	99.58%	80.09%	
IMQ31.2	At1g77740	W000032887	115.30%	99.58%	80.09%	
IMQ32.1	At1g78100	W000060346	116.83%	90.96%	88.40%	
IMQ32.2	At1g78110	W000060346	116.83%	90.96%	88.40%	
IMQ32.3	At1g78120	W000060346	116.83%	90.96%	88.40%	
IMQ33.1	At2g17036	W000176513	138.95%	98.90%	78.13%	
IMQ33.2	At2g17040	W000176513	138.95%	98.90%	78.13%	
IMQ34.1	At2g31460	W000137133	135.45%	89.55%	82.65%	
IMQ34.2	At2g31470	W000137133	135.45%	89.55%	82.65%	

Table 2

1. Gene alias	2. Tair	3. Nucleic Acid seq. GI#	4. SEQ ID NO	5. Polypeptide seq. GI#	6. SEQ ID NO	7. Putative biochemical function/ protein name	8. Conserved protein domain
IMQ21.1	At1g05230	gi 30679180	SEQ ID NO: 1	gi 30679181	SEQ ID NO: 2	DNA binding / transcription factor	IPR002913 Lipid-binding START; IPR001356 Homeobox
IMQ21.1	At1g05230	gi 30679175	SEQ ID NO: 3	gi 15220448	SEQ ID NO: 4	DNA binding / transcription factor	IPR002913 Lipid-binding START; IPR001356 Homeobox
IMQ21.2	At1g05240	gi 30679186	SEQ ID NO: 5	gi 18390498	SEQ ID NO: 6	peroxidase	IPR002016 Haem peroxidase, plant/fungal/bacterial ; IPR000823 Plant peroxidase

1. Gene alias	2. Tair	3. Nucleic Acid seq. GI#	4. SEQ ID NO	5. Polypeptide seq. GI#	6. SEQ ID NO	7. Putative biochemical function/ protein name	8. Conserved protein domain
IMQ21.3	At1g05250	gi 30679195	SEQ ID NO: 7	gi 18390500	SEQ ID NO: 8	peroxidase	IPR002016 Haem peroxidase, plant/fungal/bacterial ; IPR000823 Plant peroxidase
IMQ21.4	At1g05260	gi 30679199	SEQ ID NO: 9	gi 15220463	SEQ ID NO: 10	RCI3 (RARE COLD INDUCIBLE GENE 3); peroxidase	IPR002016 Haem peroxidase, plant/fungal/bacterial ; IPR000823 Plant peroxidase
IMQ22.1	At1g10140	gi 42561890	SEQ ID NO: 11	gi 18391115	SEQ ID NO: 12	unknown protein	
IMQ22.2	At1g10150	gi 42561891	SEQ ID NO: 13	gi 18391117	SEQ ID NO: 14	ATPP2-A10	
IMQ22.3	At1g10155	gi 22329463	SEQ ID NO: 15	gi 22329464	SEQ ID NO: 16	unknown protein	
IMQ23.1	At1g13630	gi 18394018	SEQ ID NO: 17	gi 15222912	SEQ ID NO: 18	unknown protein	IPR002885 Pentatricopeptide repeat
IMQ23.2	At1g13640	gi 30683391	SEQ ID NO: 19	gi 18394020	SEQ ID NO: 20	inositol or phosphatidylinositol kinase/ phosphotransferase, alcohol group as acceptor	IPR000403 Phosphatidylinositol 3- and 4-kinase, catalytic; IPR000626 Ubiquitin
IMQ23.3	At1g13650	gi 30683401	SEQ ID NO: 21	gi 30683402	SEQ ID NO: 22	unknown protein	
IMQ23.3	At1g13650	gi 30683399	SEQ ID NO: 23	gi 15222916	SEQ ID NO: 24	unknown protein	
IMQ24.1	At1g25400	gi 30689203	SEQ ID NO: 25	gi 18395663	SEQ ID NO: 26	unknown protein	
IMQ24.2	At1g25410	gi 18395666	SEQ ID NO: 27	gi 15222583	SEQ ID NO: 28	ATIPT6; ATP binding / adenylate dimethylallyltransferase/ tRNA isopentenyltransferase	IPR002648 Isopentenyl transferase; IPR002627 tRNA isopentenyltransferase; IPR011593 Isopentenyl transferase-like
IMQ24.3	At1g25420	gi 42571640	SEQ ID NO: 29	gi 42571641	SEQ ID NO: 30	expressed protein	IPR005061 Protein of unknown function DUF292, eukaryotic

1. Gene alias	2. Tair	3. Nucleic Acid seq. GI#	4. SEQ ID NO	5. Polypeptide seq. GI#	6. SEQ ID NO	7. Putative biochemical function/ protein name	8. Conserved protein domain
IMQ24.3	At1g25420	gi 42571638	SEQ ID NO: 31	gi 42571639	SEQ ID NO: 32	expressed protein	IPR005061 Protein of unknown function DUF292, eukaryotic
IMQ24.3	At1g25420	gi 30689214	SEQ ID NO: 33	gi 18395668	SEQ ID NO: 34	expressed protein	IPR005061 Protein of unknown function DUF292, eukaryotic
IMQ25.1	At1g27630	gi 30690014	SEQ ID NO: 35	gi 15217663	SEQ ID NO: 36	cyclin-dependent protein kinase	IPR005258 Cyclin ccl1; IPR006671 Cyclin, N-terminal
IMQ25.2	At1g27640	gi 18396427	SEQ ID NO: 37	gi 15217664	SEQ ID NO: 38	unknown protein	
IMQ25.3	At1g27650	gi 30690022	SEQ ID NO: 39	gi 15217666	SEQ ID NO: 40	RNA binding / nucleic acid binding	IPR000571 Zinc finger, CCCH-type; IPR000504 RNA-binding region RNP-1 (RNA recognition motif); IPR009145 U2 auxiliary factor small subunit
IMQ25.4	At1g27660	gi 42562353	SEQ ID NO: 41	gi 15217667	SEQ ID NO: 42	transcription factor	IPR001092 Basic helix-loop-helix dimerisation region bHLH
IMQ26.1	At1g34160	gi 18399159	SEQ ID NO: 43	gi 15218513	SEQ ID NO: 44	unknown protein	IPR002885 Pentatricopeptide repeat
IMQ26.2	At1g34170	gi 79356538	SEQ ID NO: 45	gi 79356539	SEQ ID NO: 46	ARF13; transcription factor	IPR010525 Auxin response factor; IPR003340 Transcriptional factor B3
IMQ26.2	At1g34170	gi 79319168	SEQ ID NO: 47	gi 79319169	SEQ ID NO: 48	ARF13	IPR010525 Auxin response factor; IPR003340 Transcriptional factor B3
IMQ26.3	At1g34180	gi 30693015	SEQ ID NO: 49	gi 18399166	SEQ ID NO: 50	ANAC016; transcription factor	IPR003441 No apical meristem (NAM) protein
IMQ27.1	At1g45160	gi 30693922	SEQ ID NO: 51	gi 15219539	SEQ ID NO: 52	kinase	IPR000719 Protein kinase; IPR008271 Serine/threonine protein kinase, active site; IPR000687 Protein of unknown function RIO1
IMQ27.2	At1g45170	gi 79360074	SEQ ID NO: 53	gi 79360075	SEQ ID NO: 54	unknown protein	

1. Gene alias	2. Tair	3. Nucleic Acid seq. GI#	4. SEQ ID NO	5. Polypeptide seq. GI#	6. SEQ ID NO	7. Putative biochemical function/ protein name	8. Conserved protein domain
IMQ27.3	At1g45180	gi 30693927	SEQ ID NO: 55	gi 15220067	SEQ ID NO: 56	protein binding / ubiquitin-protein ligase/ zinc ion binding	IPR001841 Zinc finger, RING-type; IPR001965 Zinc finger, PHD-type
IMQ28.1	At1g52140	gi 30695146	SEQ ID NO: 57	gi 18403871	SEQ ID NO: 58	unknown protein	
IMQ28.2	At1g52150	gi 30695148	SEQ ID NO: 59	gi 30695149	SEQ ID NO: 60	ATHB-15 (INCURVATA 4); DNA binding / transcription factor	IPR002913 Lipid-binding START; IPR001356 Homeobox; IPR004827 Basic-leucine zipper (bZIP) transcription factor
IMQ28.2	At1g52150	gi 30695147	SEQ ID NO: 61	gi 15218158	SEQ ID NO: 62	ATHB-15 (INCURVATA 4); DNA binding / transcription factor	IPR002913 Lipid-binding START; IPR001356 Homeobox; IPR004827 Basic-leucine zipper (bZIP) transcription factor
IMQ29.1	At1g58410	gi 18406289	SEQ ID NO: 63	gi 15217959	SEQ ID NO: 64	ATP binding	IPR000767 Disease resistance protein; IPR001611 Leucine-rich repeat; IPR002182 NB-ARC; IPR011072 Protein kinase PKN/PRK1, effector
IMQ29.1	At1g58410	gi 79583692	SEQ ID NO: 65	gi 79583693	SEQ ID NO: 66	ATP binding	IPR000767 Disease resistance protein; IPR001611 Leucine-rich repeat; IPR002182 NB-ARC
IMQ29.1	At1g58410	gi 79320239	SEQ ID NO: 67	gi 79320240	SEQ ID NO: 68	ATP binding	IPR000767 Disease resistance protein; IPR001611 Leucine-rich repeat; IPR002182 NB-ARC
IMQ29.2	At1g58420	gi 30696259	SEQ ID NO: 69	gi 18406291	SEQ ID NO: 70	unknown protein	
IMQ29.3	At1g58430	gi 30696261	SEQ ID NO: 71	gi 15217963	SEQ ID NO: 72	RXF26; carboxylic ester hydrolase/ hydrolase, acting on ester bonds	IPR001087 Lipolytic enzyme, G-D-S-L
IMQ30.1	At1g75670	gi 42572114	SEQ ID NO: 73	gi 42572115	SEQ ID NO: 74	unknown protein	IPR007830 RNA polymerase Rpa43 subunit
IMQ30.1	At1g75670	gi 30699111	SEQ ID NO: 75	gi 18410907	SEQ ID NO: 76	unknown protein	IPR007830 RNA polymerase Rpa43 subunit
IMQ30.2	At1g75680	gi 30699112	SEQ ID NO: 77	gi 15222328	SEQ ID NO: 78	hydrolase, hydrolyzing O-glycosyl compounds	IPR001701 Glycoside hydrolase, family 9

1. Gene alias	2. Tair	3. Nucleic Acid seq. GI#	4. SEQ ID NO	5. Polypeptide seq. GI#	6. SEQ ID NO	7. Putative biochemical function/ protein name	8. Conserved protein domain
			77		78		
IMQ30.3	At1g75690	gi 30699113	SEQ ID NO: 79	gi 15222330	SEQ ID NO: 80	unknown protein	IPR001305 DnaJ central region
IMQ30.4	At1g75700	gi 18410915	SEQ ID NO: 81	gi 15222332	SEQ ID NO: 82	unknown protein	IPR004345 TB2/DP1 and HVA22 related protein; IPR005296 IBV 3C protein
IMQ31.1	At1g77730	gi 18411662	SEQ ID NO: 83	gi 15217449	SEQ ID NO: 84	unknown protein	IPR001849 Pleckstrin-like; IPR000648 Oxysterol-binding protein
IMQ31.2	At1g77740	gi 18411668	SEQ ID NO: 85	gi 15217451	SEQ ID NO: 86	1-phosphatidylinositol-4-phosphate 5-kinase	IPR002498 Phosphatidylinositol-4-phosphate 5-kinase; IPR003409 MORN motif
IMQ32.1	At1g78100	gi 30699306	SEQ ID NO: 87	gi 18411823	SEQ ID NO: 88	unknown protein	IPR001810 Cyclin-like F-box
IMQ32.2	At1g78110	gi 42563307	SEQ ID NO: 89	gi 15218227	SEQ ID NO: 90	unknown protein	
IMQ32.3	At1g78120	gi 18411834	SEQ ID NO: 91	gi 15218228	SEQ ID NO: 92	unknown protein	IPR001440 TPR repeat; IPR005687 Mitochondrial import translocase, subunit Tom70
IMQ33.1	At2g17036	gi 18398290	SEQ ID NO: 93	gi 18398291	SEQ ID NO: 94	unknown protein	IPR001810 Cyclin-like F-box; IPR005174 Protein of unknown function DUF295
IMQ33.2	At2g17040	gi 30679858	SEQ ID NO: 95	gi 18398293	SEQ ID NO: 96	ANAC036; transcription factor	IPR003441 No apical meristem (NAM) protein
IMQ34.1	At2g31460	gi 18402675	SEQ ID NO: 97	gi 15225088	SEQ ID NO: 98	unknown protein	IPR005508 Protein of unknown function DUF313
IMQ34.2	At2g31470	gi 18402677	SEQ ID NO: 99	gi 15225089	SEQ ID NO: 100	unknown protein	IPR006527 F-box protein interaction domain; IPR001810 Cyclin-like F-box

Table 3.

1. Gene alias	2. Tair	3. Nucleic Acid seq. GI#	4. Polypeptide seq. GI#	5. Orthologous Genes: Nucleic Acid/Polypeptide seq. GI#		
				Nucleic Acid GI#	Polypeptide GI#	Species
IMQ21.1	At1g05230	gi 30679180	gi 30679181	gi 30679175	gi 15220448	Arabidopsis thaliana
				gi 42567019	gi 22328861	Arabidopsis thaliana
				gi 1173621	gi 1173622	Phalaenopsis sp. SM9108
				gi 50928892	gi 50928893	Oryza sativa (japonica cultivar-group)
IMQ21.1	At1g05230	gi 30679175	gi 15220448	gi 30679180	gi 30679181	Arabidopsis thaliana
				gi 42567019	gi 22328861	Arabidopsis thaliana
				gi 1173621	gi 1173622	Phalaenopsis sp. SM9108
				gi 50928892	gi 50928893	Oryza sativa (japonica cultivar-group)
IMQ21.2	At1g05240	gi 30679186	gi 18390498	gi 30679195	gi 18390500	Arabidopsis thaliana
				gi 30685217	gi 15242237	Arabidopsis thaliana
				gi 30678297	gi 15232058	Arabidopsis thaliana
				gi 7259218	gi 7259219	Spinacia oleracea
IMQ21.3	At1g05250	gi 30679195	gi 18390500	gi 30679186	gi 18390498	Arabidopsis thaliana
				gi 30685217	gi 15242237	Arabidopsis thaliana
				gi 30678297	gi 15232058	Arabidopsis thaliana
				gi 7259218	gi 7259219	Spinacia oleracea
IMQ21.4	At1g05260	gi 30679199	gi 15220463	gi 50261254	gi 50261255	Capsella bursa-pastoris
				gi 30686383	gi 15233153	Arabidopsis thaliana
				gi 30681721	gi 15237128	Arabidopsis thaliana
IMQ22.1	At1g10140	gi 42561890	gi 18391115	gi 30696259	gi 18406291	Arabidopsis thaliana
				gi 50933830	gi 50933831	Oryza sativa (japonica cultivar-group)
				gi 45860990	gi 50872457	Oryza sativa (japonica cultivar-group)
IMQ22.2	At1g10150	gi 42561891	gi 18391117	gi 30696298	gi 18406365	Arabidopsis thaliana
				gi 50933834	gi 50933835	Oryza sativa (japonica cultivar-group)
				gi 37514985	gi 40539064	Oryza sativa (japonica cultivar-group)
IMQ22.3	At1g10155	gi 22329463	gi 22329464	gi 18397840	gi 15221633	Arabidopsis thaliana
				gi 6850933	gi 6850934	Cicer arietinum
				gi 4995204	gi 4995205	Glycine max
IMQ23.1	At1g13630	gi 18394018	gi 15222912	gi 37535403	gi 37535404	Oryza sativa (japonica cultivar-group)
				gi 18403403	gi 15228763	Arabidopsis thaliana
				gi 18391375	gi 15221282	Arabidopsis thaliana
IMQ23.2	At1g13640	gi 30683391	gi 18394020	gi 30678087	gi 18395629	Arabidopsis thaliana
				gi 30689401	gi 18395825	Arabidopsis thaliana
				gi 42570678	gi 42570679	Arabidopsis thaliana
IMQ23.3	At1g13650	gi 30683401	gi 30683402	gi 30683399	gi 15222916	Arabidopsis thaliana
				gi 30678078	gi 30678079	Arabidopsis thaliana
				gi 42570676	gi 42570677	Arabidopsis thaliana
				gi 23496321	gi 23496344	Plasmodium falciparum 3D7
IMQ23.3	At1g13650	gi 30683399	gi 15222916	gi 30683401	gi 30683402	Arabidopsis thaliana
				gi 30678078	gi 30678079	Arabidopsis thaliana
				gi 42570676	gi 42570677	Arabidopsis thaliana
				gi 23496321	gi 23496344	Plasmodium falciparum 3D7
IMQ24.1	At1g25400	gi 30689203	gi 18395663	gi 30697678	gi 18409031	Arabidopsis thaliana
IMQ24.2	At1g25410	gi 18395666	gi 15222583	gi 18409036	gi 15221410	Arabidopsis thaliana
				gi 74038586	gi 74038587	Brassica rapa subsp. pekinensis
				gi 18402143	gi 15230294	Arabidopsis thaliana
IMQ24.3	At1g25420	gi 42571640	gi 42571641	gi 42571638	gi 42571639	Arabidopsis thaliana
				gi 30689214	gi 18395668	Arabidopsis thaliana
				gi 34911297	gi 34911298	Oryza sativa (japonica cultivar-group)
				gi 16191725	gi 56784451	Oryza sativa (japonica cultivar-group)
IMQ24.3	At1g25420	gi 42571638	gi 42571639	gi 42571640	gi 42571641	Arabidopsis thaliana
				gi 30689214	gi 18395668	Arabidopsis thaliana

1. Gene alias	2. Tair	3. Nucleic Acid seq. GI#	4. Polypeptide seq. GI#	5. Orthologous Genes: Nucleic Acid/Polypeptide seq. GI#		
				Nucleic Acid GI#	Polypeptide GI#	Species
				gi 34911297	gi 34911298	Oryza sativa (japonica cultivar-group)
				gi 16191725	gi 56784451	Oryza sativa (japonica cultivar-group)
IMQ24.3	At1g25420	gi 30689214	gi 18395668	gi 42571640	gi 42571641	Arabidopsis thaliana
				gi 42571638	gi 42571639	Arabidopsis thaliana
				gi 16191725	gi 56784451	Oryza sativa (japonica cultivar-group)
				gi 42571740	gi 42571741	Arabidopsis thaliana
IMQ25.1	At1g27630	gi 30690014	gi 15217663	gi 30694714	gi 30694715	Arabidopsis thaliana
				gi 77548247	gi 77548754	Oryza sativa (japonica cultivar-group)
				gi 30684821	gi 30684822	Arabidopsis thaliana
IMQ25.2	At1g27640	gi 18396427	gi 15217664	gi 214833	gi 214834	Xenopus laevis
IMQ25.3	At1g27650	gi 30690022	gi 15217666	gi 30694150	gi 15239067	Arabidopsis thaliana
				gi 42573546	gi 42573547	Arabidopsis thaliana
				gi 13278054	gi 13278055	Mus musculus
				gi 3850815	gi 3850816	Oryza sativa
IMQ25.4	At1g27660	gi 42562353	gi 15217667	gi 55769700	gi 55769701	Oryza sativa (japonica cultivar-group)
				gi 58532108	gi 21741062	Oryza sativa (japonica cultivar-group)
				gi 30696593	gi 18407276	Arabidopsis thaliana
IMQ26.1	At1g34160	gi 18399159	gi 15218513	gi 77552765	gi 77554579	Oryza sativa (japonica cultivar-group)
				gi 54695179	gi 54695180	Physcomitrella patens
				gi 45935109	gi 45935146	Ipomoea trifida
IMQ26.2	At1g34170	gi 79356538	gi 79356539	gi 42562509	gi 42562510	Arabidopsis thaliana
				gi 18399246	gi 15218610	Arabidopsis thaliana
				gi 18399735	gi 15219635	Arabidopsis thaliana
IMQ26.2	At1g34170	gi 79319168	gi 79319169	gi 42562509	gi 42562510	Arabidopsis thaliana
				gi 18399735	gi 15219635	Arabidopsis thaliana
				gi 18399246	gi 15218610	Arabidopsis thaliana
IMQ26.3	At1g34180	gi 30693015	gi 18399166	gi 30693016	gi 18399168	Arabidopsis thaliana
				gi 21105741	gi 21105742	Petunia x hybrida
				gi 21105739	gi 21105740	Petunia x hybrida
IMQ27.1	At1g45160	gi 30693922	gi 15219539	gi 30684701	gi 30684702	Arabidopsis thaliana
				gi 50918242	gi 50918243	Oryza sativa (japonica cultivar-group)
				gi 30697658	gi 15241795	Arabidopsis thaliana
IMQ27.2	At1g45170	gi 79360074	gi 79360075	gi 42562572	gi 42562573	Arabidopsis thaliana
				gi 30694183	gi 18422277	Arabidopsis thaliana
				gi 2764573	gi 2764574	Pisum sativum
IMQ27.3	At1g45180	gi 30693927	gi 15220067	gi 22327533	gi 15239131	Arabidopsis thaliana
				gi 42569056	gi 15226553	Arabidopsis thaliana
				gi 42570792	gi 42570793	Arabidopsis thaliana
				gi 50929180	gi 50929181	Oryza sativa (japonica cultivar-group)
IMQ28.1	At1g52140	gi 30695146	gi 18403871	gi 30684002	gi 15228179	Arabidopsis thaliana
				gi 12003387	gi 12003388	Nicotiana tabacum
				gi 18417331	gi 15233454	Arabidopsis thaliana
IMQ28.2	At1g52150	gi 30695148	gi 30695149	gi 30695147	gi 15218158	Arabidopsis thaliana
				gi 60327628	gi 60327629	Populus trichocarpa
				gi 60327630	gi 60327631	Populus trichocarpa
IMQ28.2	At1g52150	gi 30695147	gi 15218158	gi 30695148	gi 30695149	Arabidopsis thaliana
				gi 60327628	gi 60327629	Populus trichocarpa
				gi 60327630	gi 60327631	Populus trichocarpa
IMQ29.1	At1g58410	gi 18406289	gi 15217959	gi 18406284	gi 15217957	Arabidopsis thaliana
				gi 42562806	gi 15218003	Arabidopsis thaliana
				gi 79320239	gi 79320240	Arabidopsis thaliana
				gi 79583692	gi 79583693	Arabidopsis thaliana
				gi 18406280	gi 15217954	Arabidopsis thaliana

1. Gene alias	2. Tair	3. Nucleic Acid seq. GI#	4. Polypeptide seq. GI#	5. Orthologous Genes: Nucleic Acid/Polypeptide seq. GI#		
				Nucleic Acid GI#	Polypeptide GI#	Species
IMQ29.1	At1g58410	gi 79583692	gi 79583693	gi 42562806	gi 15218003	Arabidopsis thaliana
				gi 79320239	gi 79320240	Arabidopsis thaliana
				gi 30696274	gi 22330316	Arabidopsis thaliana
				gi 22330305	gi 22330306	Arabidopsis thaliana
				gi 30696285	gi 30696286	Arabidopsis thaliana
IMQ29.1	At1g58410	gi 79320239	gi 79320240	gi 42562806	gi 15218003	Arabidopsis thaliana
				gi 79583692	gi 79583693	Arabidopsis thaliana
				gi 30696274	gi 22330316	Arabidopsis thaliana
				gi 22330305	gi 22330306	Arabidopsis thaliana
				gi 30696285	gi 30696286	Arabidopsis thaliana
IMQ29.2	At1g58420	gi 30696259	gi 18406291	gi 42561890	gi 18391115	Arabidopsis thaliana
				gi 50933830	gi 50933831	Oryza sativa (japonica cultivar-group)
				gi 30698165	gi 30698166	Arabidopsis thaliana
IMQ29.3	At1g58430	gi 30696261	gi 15217963	gi 18402698	gi 15225096	Arabidopsis thaliana
				gi 18402315	gi 15227734	Arabidopsis thaliana
				gi 18402293	gi 15227723	Arabidopsis thaliana
IMQ30.1	At1g75670	gi 42572114	gi 42572115	gi 30699111	gi 18410907	Arabidopsis thaliana
				gi 55741413	gi 55741414	Oryza sativa (japonica cultivar-group)
				gi 34898425	gi 34898426	Oryza sativa (japonica cultivar-group)
				gi 66828550	gi 66828551	Dictyostelium discoideum
IMQ30.1	At1g75670	gi 30699111	gi 18410907	gi 42572114	gi 42572115	Arabidopsis thaliana
				gi 55741413	gi 55741414	Oryza sativa (japonica cultivar-group)
				gi 34898425	gi 34898426	Oryza sativa (japonica cultivar-group)
				gi 66828550	gi 66828551	Dictyostelium discoideum
IMQ30.2	At1g75680	gi 30699112	gi 15222328	gi 18394803	gi 15223718	Arabidopsis thaliana
				gi 50725787	gi 50725801	Oryza sativa (japonica cultivar-group)
				gi 46849602	gi 53791781	Oryza sativa (japonica cultivar-group)
IMQ30.3	At1g75690	gi 30699113	gi 15222330	gi 4732090	gi 4732091	Zea mays
				gi 68433925	gi 68433926	Danio rerio
				gi 13430173	gi 13430174	Castanea sativa
IMQ30.4	At1g75700	gi 18410915	gi 15222332	gi 18394804	gi 18394805	Arabidopsis thaliana
				gi 30694082	gi 18422223	Arabidopsis thaliana
				gi 50919650	gi 50919651	Oryza sativa (japonica cultivar-group)
IMQ31.1	At1g77730	gi 18411662	gi 15217449	gi 42572842	gi 42572843	Arabidopsis thaliana
				gi 42570130	gi 42570131	Arabidopsis thaliana
				gi 30680661	gi 30680662	Arabidopsis thaliana
IMQ31.2	At1g77740	gi 18411668	gi 15217451	gi 30687626	gi 15219152	Arabidopsis thaliana
				gi 50918122	gi 50918123	Oryza sativa (japonica cultivar-group)
				gi 8885991	gi 8885992	Nicotiana rustica
IMQ32.1	At1g78100	gi 30699306	gi 18411823	gi 42562235	gi 15219845	Arabidopsis thaliana
				gi 18844754	gi 55297493	Oryza sativa (japonica cultivar-group)
				gi 50931860	gi 50931861	Oryza sativa (japonica cultivar-group)
IMQ32.2	At1g78110	gi 42563307	gi 15218227	gi 18395090	gi 15219847	Arabidopsis thaliana
				gi 50899793	gi 50899794	Oryza sativa (japonica cultivar-group)
				gi 50912868	gi 50912869	Oryza sativa (japonica cultivar-group)
IMQ32.3	At1g78120	gi 18411834	gi 15218228	gi 42568787	gi 15238361	Arabidopsis thaliana
				gi 18416178	gi 15238058	Arabidopsis thaliana
				gi 37694873	gi 72255609	Brassica rapa
IMQ33.1	At2g17036	gi 18398290	gi 18398291	gi 42569091	gi 18398287	Arabidopsis thaliana
				gi 18424314	gi 15238601	Arabidopsis thaliana
				gi 42562951	gi 42562952	Arabidopsis thaliana
IMQ33.2	At2g17040	gi 30679858	gi 18398293	gi 66394519	gi 66394520	Glycine max
				gi 54291125	gi 54291129	Oryza sativa (japonica cultivar-group)

1. Gene alias	2. Tair	3. Nucleic Acid seq. GI#	4. Polypeptide seq. GI#	5. Orthologous Genes: Nucleic Acid/Polypeptide seq. GI#		
				Nucleic Acid GI#	Polypeptide GI#	Species
				gi 30678001	gi 30678002	Arabidopsis thaliana
IMQ34.1	At2g31460	gi 18402675	gi 15225088	gi 18401417	gi 15225878	Arabidopsis thaliana
				gi 18402662	gi 15224674	Arabidopsis thaliana
				gi 18408901	gi 15229174	Arabidopsis thaliana
IMQ34.2	At2g31470	gi 18402677	gi 15225089	gi 18403984	gi 15229553	Arabidopsis thaliana
				gi 18424999	gi 15239182	Arabidopsis thaliana
				gi 18408857	gi 15229145	Arabidopsis thaliana

EXAMPLE 2

Analysis of the *Arabidopsis* IMQ Sequence

- 5 Sequence analyses were performed with BLAST (Altschul *et al.*, 1990, *J. Mol. Biol.* 215:403-410), PFAM (Bateman *et al.*, 1999, *Nucleic Acids Res.* 27:260-262), INTERPRO (Mulder *et al.* 2003 *Nucleic Acids Res.* 31, 315-318.), 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). Conserved domains
- 10 for each protein are listed in column 8 of Table 2.

EXAMPLE 3

- To test whether over-expression of the genes identified in Tables 1-3 alter the seed composition phenotype, protein, digestible protein, oil and fiber content in
- 15 seeds from transgenic plants expressing these genes was compared with protein, digestible protein, oil and fiber content in seeds from non-transgenic control plants. To do this, the genes were cloned into plant transformation vectors behind the strong constitutive CsVMV promoter and the seed specific PRU promoter. These constructs were transformed into *Arabidopsis* plants using the floral dip method.
- 20 The plant transformation vector contains a gene, which provides resistance to a toxic compound, and serves as a selectable marker. Seed from the transformed plants were plated on agar medium containing the toxic compound. After 7 days, transgenic plants were identified as healthy green plants and transplanted to soil. Non-transgenic control plants were germinated on agar medium, allowed to grow for
- 25 7 days and then transplanted to soil. Transgenic seedlings and non-transgenic control plants were transplanted to two inch pots that were placed in random

positions in a 10 inch by 20 inch tray. The plants were grown to maturity, allowed to self-fertilize and set seed. Seed was harvested from each plant and its oil content estimated by Near Infrared (NIR) Spectroscopy using methods previously described. The effect of each construct on seed composition was examined in at least two experiments.

Table 4 lists constructs tested for causing a significant increase in oil, protein, digestible protein or a significant decrease in fiber were identified by a two-way Analysis of Variance (ANOVA) test at a p-value ≤ 0.05 . The ANOVA p-values for Protein, Oil, Digestible Protein and Fiber are listed in columns 4-7, respectively. Those with a significant p-value are listed in bold. The Average values for Protein, Oil, Digestible Protein and Fiber are listed in columns 8-11, respectively and were calculated by averaging the average values determined for the transgenic plants in each experiment.

Table 4.

1. Gene	2. TAIR	3. Construct	4. ANOVA Protein	5. ANOVA Oil	6. ANOVA Digestible Protein	7. ANOVA Fiber	8. Protein	9. Oil	10. Digestible Protein	11. Fiber
IMQ25.3	At1g27650	CsVMV::At1g27650	0.000	0.052	0.010	0.003	89.8%	103.7%	95.9%	105.1%
IMQ25.3	At1g27650	Pru::At1g27650	0.004	0.126	0.024	0.266	94.0%	102.4%	98.0%	100.9%
IMQ29.1	At1g58410	CsVMV::At1g58410	0.012	0.005	0.279	0.775	104.4%	96.7%	101.2%	99.7%
IMQ29.1	At1g58410	Pru::At1g58410	0.993	0.805	0.090	0.025	100.1%	99.6%	101.5%	97.9%
IMQ29.2	At1g58420	CsVMV::At1g58420	0.000	0.011	0.689	0.145	95.93%	103.73%	99.60%	98.73%
IMQ30.3	At1g75690	Pru::At1g75690	0.051	0.007	0.851	0.326	96.0%	104.9%	99.9%	99.0%

EXAMPLE 4

To test whether over-expression of the genes identified in Tables 1-4 alter the seed composition phenotype, protein, digestible protein, oil, and fiber content in seeds from transgenic plants expressing these genes is compared with protein, digestible protein, oil and fiber content in seeds from non-transgenic control plants. Any one of the genes identified in Tables 1-4 is used to transform *Brassica napus* (canola). To do this, the genes are cloned into plant transformation vectors behind the strong constitutive CsVMV promoter and the seed specific phaseolin promoter. These constructs (which include a gene encoding a selection agent) are transformed into canola plants.

Transformation of canola is accomplished via *Agrobacterium*-mediated transformation. Seeds are surface-sterilized with 10% commercial bleach for 10 minutes and rinsed 3 times with sterile distilled water. The seeds are then placed on one half concentration of MS basal medium (Murashige and Skoog, *Physiol. Plant.* 15:473-497, 1962) and maintained under growth regime set at 25° C, and a photoperiod of 16 hrs light/8 hrs dark.

Hypocotyl segments (3-5 mm) are excised from 5 - 7 day old seedlings and placed on callus induction medium K1D1 (MS medium with 1 mg/l kinetin and 1 mg/l 2,4-D) for 3 days as pre-treatment. The segments are then transferred into a petri plate, treated with *Agrobacterium* Z707S or LBA4404 strain containing pDAB721. The *Agrobacterium* is grown overnight at 28° C in the dark on a shaker at 150 rpm and subsequently re-suspended in the culture medium.

After 30 minute treatment of the hypocotyl segments with *Agrobacterium*, these are placed back on the callus induction medium for 3 days. Following co-cultivation, the segments are placed on K1D1TC (callus induction medium containing 250 mg/l Carbenicillin and 300 mg/l Timentin) for one week of recovery. Alternately, the segments are placed directly on selection medium K1D1H1 (above medium with 1 mg/l selection agent, for example an herbicide). Carbenicillin and Timentin are antibiotics used to kill the *Agrobacterium*. The selection agent is used to allow the growth of the transformed cells.

Callus samples from independent events are tested by PCR. All the samples tested are positive for the presence of the transformed gene, whereas the non-transformed controls are negative. Callus samples are confirmed to express the appropriate protein as determined by ELISA.

Callused hypocotyl segments are then placed on B3Z1H1 (MS medium, 3 mg/l benzylamino purine, 1 mg/l Zeatin, 0.5 gm/l MES [2-(N-morpholino) ethane sulfonic acid], 5 mg/l silver nitrate, 1 mg/l selection agent, Carbenicillin and Timentin) shoot regeneration medium. After shoots start to regenerate (approximately 3 weeks), hypocotyl segments along with the shoots are transferred to B3Z1H3 medium (MS medium, 3 mg/l benzylamino purine, 1 mg/l Zeatin, 0.5 gm/l MES [2-(N-morpholino) ethane sulfonic acid], 5 mg/l silver nitrate, 3 mg/l selection agent, Carbenicillin and Timentin) for 3 weeks.

Shoots are excised from the hypocotyl segments and transferred to shoot elongation medium MESH10 (MS, 0.5 gm/l MES, 10 mg/l selection agent, Carbenicillin, Timentin) for 2-4 weeks. The elongated shoots are cultured for root induction on MSI.1 (MS with 0.1 mg/l Indolebutyric acid). Once the plants have a well established root system, these are transplanted into soil. The plants are acclimated under controlled environmental conditions in the Conviron for 1-2 weeks before transfer to the greenhouse. The transformed T0 plants self-pollinate in the greenhouse to obtain T1 seed. Transgenic plants are selected at the T1 generation based on resistance to a selection agent. T2 seed (from T1 plants) is harvested and sown in soil. T2 plants are grown to maturity, allowed to self-fertilize and set seed. T3 seed (from the T2 plants) is harvested in bulk for each line. Seed oil, protein, digestible protein, and fiber values are measured as discussed in Example 1.

IT IS CLAIMED:

1. A method of producing an improved seed quality 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 an IMQ polypeptide comprising an amino acid sequence at least 95% identical to the full length of the amino acid sequence set forth in SEQ ID NO: 80,

b) growing the transformed progenitor cells to produce a transgenic plant, wherein the nucleotide sequence is expressed, and the transgenic plant exhibits an improved seed quality phenotype relative to control plants, and wherein the improved seed quality phenotype comprises an increase in oil content, and

c) measuring the improved seed quality phenotype in the transgenic plant, relative to control plants,
thereby producing the improved seed quality phenotype in the plant.

2. The method of claim 1, wherein the plant is of a *Brassica* species.

3. The method of claim 1, wherein the plant is a canola, rapeseed, soy, corn, sunflower, cotton, cocoa, safflower, oil palm, coconut palm, flax, castor, peanut, wheat, oat, or rice plant.

4. The method of claim 3, wherein the plant is a canola plant.

5. The method of any one of claims 1 to 4, wherein the IMQ polypeptide comprises the amino acid sequence set forth in SEQ ID NO: 80.

6. The method of any one of claims 1 to 4, wherein the IMQ polypeptide consists of the amino acid sequence set forth in SEQ ID NO: 80.

7. The method of any one of claims 1 to 4, wherein the nucleotide sequence comprises the nucleotide sequence set forth as SEQ ID NO: 79.

8. The method of any one of claims 1 to 6, wherein the nucleotide sequence is operably linked to a constitutive, inducible, or regulatable promoter sequence.
9. A method of generating a plant having an improved seed quality phenotype comprising
 identifying a plant that has an allele in an IMQ gene, where the wild-type IMQ gene has the nucleotide sequence set forth as SEQ ID NO: 79, which allele results in improved seed quality phenotype compared to a control plant lacking the allele;
 generating progeny of said identified plant, wherein the generated progeny inherit the allele and have the improved seed quality phenotype; and
 measuring the improved seed quality phenotype in the progeny comprising the allele, relative to the control plant lacking the allele,
thereby generating a plant having an improved seed quality phenotype.
10. The method of claim 9 that employs candidate gene/QTL methodology to identify the plant.
11. The method of claim 9 that employs TILLING methodology to identify the plant.