

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
15 December 2005 (15.12.2005)

PCT

(10) International Publication Number  
**WO 2005/118821 A2**

- (51) International Patent Classification<sup>7</sup>: **C12N 15/82**
- (21) International Application Number:  
PCT/US2005/018918
- (22) International Filing Date: 26 May 2005 (26.05.2005)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/575,561 28 May 2004 (28.05.2004) US
- (71) Applicant (for all designated States except US): **AGRI-NOMICS LLC** [US/US]; 16160 SW Upper Boones Ferry Road, Portland, OR 99224-7744 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **DAVIES, John** [US/US]; 4617 S.E. Madison Street, Portland, OR 97215 (US). **PETERS, Sandra** [US/US]; 5029 S.E. 40th Avenue, Portland, OR 97202 (US). **NG, Hein, Tsoeng** [—/US]; 12975 SW 19th Street, Beaverton, OR 97008 (US).
- (74) Agent: **CARLSON, Anne**; Klarquist Sparkman, LLP, One World Trade Center, Suite 1600, 121 SW Salmon Street, Portland, OR 97204 (US).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Declaration under Rule 4.17:**  
— of inventorship (Rule 4.17(iv)) for US only
- Published:**  
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



**WO 2005/118821 A2**

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

(57) Abstract: The present invention is directed to plants that display an altered oil content phenotype due to altered expression of a HIO1005 nucleic acid. The invention further directed to methods of generating plants with an altered oil content phenotype.

## GENERATION OF PLANTS WITH ALTERED OIL CONTENT

### REFERENCE TO RELATED APPLICATIONS

This application claims priority to U.S. provisional application 60/575,561 filed  
5 May 28, 2004, the contents of which are hereby incorporated by reference.

### BACKGROUND OF THE INVENTION

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

Biotechnological manipulation of oils can provide compositional alteration and  
improvement of oil yield. Compositional alterations include high oleic soybean and corn oil  
(US Pat Nos 6,229,033 and 6,248,939), and laurate-containing seeds (US Pat No 5,639,790),  
among others. Work in compositional alteration has predominantly focused on processed  
35 oilseeds but has been readily extendable to non-oilseed crops, including corn. While there is

considerable interest in increasing oil content, the only currently practiced biotechnology in this area is High-Oil Corn (HOC) technology (DuPont, US PAT NO: 5,704,160). HOC employs high oil pollinators developed by classical selection breeding along with elite (male-sterile) hybrid females in a production system referred to as TopCross. The TopCross  
5 High Oil system raises harvested grain oil content in maize from about 3.5% to about 7%, improving the energy content of the grain.

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

15 The most obvious target crops for the processed oil market are soy and rapeseed, and a large body of commercial work (e.g., US Pat No: 5,952,544; PCT application WO9411516) demonstrates that *Arabidopsis* is an excellent model for oil metabolism in these crops. Biochemical screens of seed oil composition have identified *Arabidopsis* genes for many critical biosynthetic enzymes and have led to identification of agronomically  
20 important gene orthologs. For instance, screens using chemically mutagenized populations have identified lipid mutants whose seeds display altered fatty acid composition (Lemieux *et al.*, 1990; James and Dooner, 1990). T-DNA mutagenesis screens (Feldmann *et al.*, 1989) that detected altered fatty acid composition identified the omega 3 desaturase (*FAD3*) and delta-12 desaturase (*FAD2*) genes (US Pat No 5952544; Yadav *et al.*, 1993; Okuley *et al.*,  
25 1994). A screen which focused on oil content rather than oil quality, analyzed chemically-induced mutants for wrinkled seeds or altered seed density, from which altered seed oil content was inferred (Focks and Benning, 1998). Another screen, designed to identify enzymes involved in production of very long chain fatty acids, identified a mutation in the gene encoding a diacylglycerol acyltransferase (DGAT) as being responsible for reduced  
30 triacyl glycerol accumulation in seeds (Katavic V *et al.*, 1995). It was further shown that seed-specific over-expression of the DGAT cDNA was associated with increased seed oil content (Jako *et al.*, 2001).

Activation tagging in plants refers to a method of generating random mutations by insertion of a heterologous nucleic acid construct comprising regulatory sequences (e.g., an  
35 enhancer) into a plant genome. The regulatory sequences can act to enhance transcription of

one or more native plant genes; accordingly, activation tagging is a fruitful method for generating gain-of-function, generally dominant mutants (see, *e.g.*, Hayashi *et al.*, 1992; Weigel D *et al.* 2000). The inserted construct provides a molecular tag for rapid identification of the native plant whose mis-expression causes the mutant phenotype.

5 Activation tagging may also cause loss-of-function phenotypes. The insertion may result in disruption of a native plant gene, in which case the phenotype is generally recessive.

Activation tagging has been used in various species, including tobacco and *Arabidopsis*, to identify many different kinds of mutant phenotypes and the genes associated with these phenotypes (Wilson *et al.*, 1996, Schaffer *et al.*, 1998, Fridborg *et al.*, 1999; 10 Kardailsky *et al.*, 1999; Christensen S *et al.*, 1998).

#### SUMMARY OF THE INVENTION

The invention provides a transgenic plant having a high oil phenotype. The transgenic plant comprises a transformation vector comprising a nucleotide sequence that 15 encodes or is complementary to a sequence that encodes a HIO1005 polypeptide. In preferred embodiments, the transgenic plant is selected from the group consisting of rapeseed, soy, corn, sunflower, cotton, cocoa, safflower, oil palm, coconut palm, flax, castor and peanut. The invention further provides a method of producing oil comprising growing the transgenic plant and recovering oil from said plant.

20 The invention also provides a transgenic plant cell having a high oil phenotype. The transgenic plant cell comprises a transformation vector comprising a nucleotide sequence that encodes or is complementary to a sequence that encodes a High Oil (hereinafter "HIO1005") polypeptide. In preferred embodiments, the transgenic plant cell is selected from the group consisting of rapeseed, soy, corn, sunflower, cotton, cocoa, 25 safflower, oil palm, coconut palm, flax, castor and peanut. In other embodiments, the plant cell is a seed, pollen, propagule, or embryo cell. The invention further provides feed, meal, grain, food, or seed comprising a nucleic acid sequence that encodes a HIO1005 polypeptide. The invention also provides feed, meal, grain, food, or seed comprising the HIO1005 polypeptide, or an ortholog thereof.

30 The transgenic plant of the invention is produced by a method that comprises introducing into progenitor cells of the plant a plant transformation vector comprising a nucleotide sequence that encodes or is complementary to a sequence that encodes a HIO1005 polypeptide, and growing the transformed progenitor cells to produce a transgenic plant, wherein the HIO1005 polynucleotide sequence is expressed causing the high oil 35 phenotype.

The present invention also provides a container of over about 10,000, more preferably about 20,000, and even more preferably about 40,000 seeds where over about 10%, more preferably about 25%, more preferably about 50%, and even more preferably about 75% or more preferably about 90% of the seeds are seeds derived from a plant of the present invention. The invention further provides plant cells obtained from said transgenic plant.

The present invention also provides a container of over about 10 kg, more preferably about 25 kg, and even more preferably about 50 kg seeds where over about 10%, more preferably about 25%, more preferably about 50%, and even more preferably about 75% or more preferably about 90% of the seeds are seeds derived from a plant of the present invention.

Any of the plants or parts thereof of the present invention may be processed to produce a feed, meal, or oil preparation. A particularly preferred plant part for this purpose is a seed. In a preferred embodiment the feed, meal, or oil preparation is designed for ruminant animals. Methods to produce feed, meal, and oil preparations are known in the art. *See*, for example, U.S. Patents 4,957,748; 5,100,679; 5,219,596; 5,936,069; 6,005,076; 6,146,669; and 6,156,227. The meal of the present invention may be blended with other meals. In a preferred embodiment, the meal produced from plants of the present invention or generated by a method of the present invention constitutes greater than about 0.5%, about 1%, about 5%, about 10%, about 25%, about 50%, about 75%, or about 90% by volume or weight of the meal component of any product. In another embodiment, the meal preparation may be blended and can constitute greater than about 10%, about 25%, about 35%, about 50%, or about 75% of the blend by volume.

## 25 DETAILED DESCRIPTION OF THE INVENTION

### Definitions

Unless otherwise indicated, all technical and scientific terms used herein have the same meaning as they would to one skilled in the art of the present invention. Practitioners are particularly directed to Sambrook *et al.*, 1989, and Ausubel FM *et al.*, 1993, for definitions and terms of the art. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary.

As used herein, the term "vector" refers to a nucleic acid construct designed for transfer between different host cells. An "expression vector" refers to a vector that has the ability to incorporate and express heterologous DNA fragments in a foreign cell. Many

prokaryotic and eukaryotic expression vectors are commercially available. Selection of appropriate expression vectors is within the knowledge of those having skill in the art.

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

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

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

As used herein, the term "gene expression" refers to the process by which a polypeptide is produced based on the nucleic acid sequence of a gene. The process includes both transcription and translation; accordingly, "expression" may refer to either a polynucleotide or polypeptide sequence, or both. Sometimes, expression of a polynucleotide sequence will not lead to protein translation. "Over-expression" refers to increased expression of a polynucleotide and/or polypeptide sequence relative to its expression in a wild-type (or other reference [*e.g.*, non-transgenic]) plant and may relate to a naturally-occurring or non-naturally occurring sequence. "Ectopic expression" refers to expression at a time, place, and/or increased level that does not naturally occur in the non-altered or wild-type plant. "Under-expression" refers to decreased expression of a polynucleotide and/or polypeptide sequence, generally of an endogenous gene, relative to its

expression in a wild-type plant. The terms "mis-expression" and "altered expression" encompass over-expression, under-expression, and ectopic expression.

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

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

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

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

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

As used herein, the term "T1" refers to the generation of plants from the seed of T0  
35 plants. The T1 generation is the first set of transformed plants that can be selected by

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

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

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

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.

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* 5 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. 10 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 15 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 20 "selection" steps.

With respect to microprojectile bombardment (U.S. Patent No. 5,550,318 (Adams *et al.*); U.S. Patent No. 5,538,880 (Lundquist *et al.*), U.S. Patent No. 5,610,042 (Chang *et al.*); and PCT Publication WO 95/06128 (Adams *et al.*); each of which is specifically 25 incorporated herein by reference in its entirety), 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 30 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 35 microprojectile bombardment include monocot species such as maize (International

Publication No. WO 95/06128 (*Adams et al.*)), barley, wheat (U.S. Patent No. 5,563,055 (Townsend *et al.*) incorporated herein by reference in its entirety), rice, oat, rye, sugarcane, and sorghum; as well as a number of dicots including tobacco, soybean (U.S. Patent No. 5,322,783 (Tomes *et al.*), incorporated herein by reference in its entirety), sunflower, peanut, cotton, tomato, and legumes in general (U.S. Patent No. 5,563,055 (Townsend *et al.*) incorporated herein by reference in its entirety).

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 regenerable 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; kanamycin 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 (Barry, *et al.*), U.S. Patent No 5,633,435 (Barry, *et al.*), and U.S. Patent No 6,040,497 (Spencer, *et al.*) and aroA described in U.S. Patent No. 5,094,945 (Comai) for glyphosate tolerance; a polynucleotide molecule encoding bromoxynil nitrilase (Bxn) described in U.S. Patent No. 4,810,648 (Duerrschabel, *et al.*) for Bromoxynil tolerance; a polynucleotide molecule encoding phytoene desaturase (crtI) described in Misawa et al, (1993) Plant J. 4:833-840 and Misawa et al, (1994) Plant J. 6:481-489 for norflurazon tolerance; a polynucleotide molecule encoding acetohydroxyacid synthase (AHAS, aka ALS) described in Sathasiivan *et al.* (1990) Nucl. Acids Res. 18:2188-2193 for tolerance to sulfonylurea herbicides; and the bar gene described in DeBlock, *et al.* (1987) EMBO J. 6:2513-2519 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.

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 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 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 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 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 breeding techniques can be used, depending upon the species to be crossed.

#### **Identification of Plants with an Altered Oil Content Phenotype**

We used an *Arabidopsis* activation tagging screen to identify the association between the gene we have identified and designated "HIO1005," (At5g02290, GI#30679585, GI#30679589) encoding a protein kinase protein (GI# 15241749, 30679590), respectively, and an altered oil content phenotype (specifically, a high oil phenotype). Briefly, and as further described in the Examples, a large number of *Arabidopsis* plants were mutated with the pSKI015 vector, which comprises a T-DNA from the Ti plasmid of *Agrobacterium tumefaciens*, a viral enhancer element, and a selectable marker gene (Weigel

*et al.*, 2000). When the T-DNA inserts into the genome of transformed plants, the enhancer element can cause up-regulation genes in the vicinity, generally within about 10 kilobase (kb) of the insertion. T1 plants were exposed to the selective agent in order to specifically recover transformed plants that expressed the selectable marker and therefore harbored T-DNA insertions. Samples of approximately 15-20 T2 seeds were collected from transformed T1 plants, and lipids were extracted from whole seeds. Gas chromatography (GC) analysis was performed to determine fatty acid content and composition of seed samples.

An *Arabidopsis* line that showed a high-oil phenotype was identified. The association of the HIO1005 gene with the high oil phenotype was discovered by analysis of the genomic DNA sequence flanking the T-DNA insertion in the identified line. Accordingly, HIO1005 genes and/or polypeptides may be employed in the development of genetically modified plants having a modified oil content phenotype ("a HIO1005 phenotype"). HIO1005 genes may be used in the generation of oilseed crops that provide improved oil yield from oilseed processing and in the generation of feed grain crops that provide increased energy for animal feeding. HIO1005 genes may further be used to increase the oil content of specialty oil crops, in order to augment yield of desired unusual fatty acids. Transgenic plants that have been genetically modified to express HIO1005 can be used in the production of oil, wherein the transgenic plants are grown, and oil is obtained from plant parts (e.g. seed) using standard methods.

#### **HIO1005 Nucleic Acids and Polypeptides**

*Arabidopsis* HIO1005 nucleic acid sequence is provided in SEQ ID NO:1 and in Genbank entry GI#30679585. The corresponding protein sequence is provided in SEQ ID NO:3 and in GI# 15241749. One putative variant (SEQ. ID NO:2) was also identified which encodes protein (SEQ ID NO:4) that is 100% identical with SEQ ID NO:3. Nucleic acids and/or proteins that are orthologs or paralogs of *Arabidopsis* HIO1005, are described in Example 4 below.

As used herein, the term "HIO1005 polypeptide" refers to a full-length HIO1005 protein or a fragment, derivative (variant), or ortholog thereof that is "functionally active," meaning that the protein fragment, derivative, or ortholog exhibits one or more of the functional activities associated with the polypeptide of SEQ ID NO:2. In one preferred embodiment, a functionally active HIO1005 polypeptide causes an altered oil content phenotype when mis-expressed in a plant. In a further preferred embodiment, mis-expression of the HIO1005 polypeptide causes a high oil phenotype in a plant. In another

embodiment, a functionally active HIO1005 polypeptide is capable of rescuing defective (including deficient) endogenous HIO1005 activity when expressed in a plant or in plant cells; the rescuing polypeptide may be from the same or from a different species as that with defective activity. In another embodiment, a functionally active fragment of a full length HIO1005 polypeptide (i.e., a native polypeptide having the sequence of SEQ ID NO:2 or a naturally occurring ortholog thereof) retains one or more of the biological properties associated with the full-length HIO1005 polypeptide, such as signaling activity, binding activity, catalytic activity, or cellular or extra-cellular localizing activity. A HIO1005 fragment preferably comprises a HIO1005 domain, such as a C- or N-terminal or catalytic domain, among others, and preferably comprises at least 10, preferably at least 20, more preferably at least 25, and most preferably at least 50 contiguous amino acids of a HIO1005 protein. Functional domains can be identified using the PFAM program (Bateman A *et al.*, 1999 *Nucleic Acids Res* 27:260-262). A preferred HIO1005 fragment comprises of one or more protein kinase domains.

Functionally active variants of full-length HIO1005 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 HIO1005 polypeptide. In some cases, variants are generated that change the post-translational processing of a HIO1005 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 "HIO1005 nucleic acid" encompasses nucleic acids with the sequence provided in or complementary to the sequence provided in SEQ ID NO:1, as well as functionally active fragments, derivatives, or orthologs thereof. A HIO1005 nucleic acid of this invention may be DNA, derived from genomic DNA or cDNA, or RNA.

In one embodiment, a functionally active HIO1005 nucleic acid encodes or is complementary to a nucleic acid that encodes a functionally active HIO1005 polypeptide. Included within this definition is genomic DNA that serves as a template for a primary RNA transcript (i.e., an mRNA precursor) that requires processing, such as splicing, before encoding the functionally active HIO1005 polypeptide. A HIO1005 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 HIO1005 polypeptide, or an intermediate form. A HIO1005 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 HIO1005 nucleic acid is capable of being used in the generation of loss-of-function HIO1005 phenotypes, for instance, via antisense suppression, co-suppression, etc.

In one preferred embodiment, a HIO1005 nucleic acid used in the methods of this invention comprises a nucleic acid sequence that encodes or is complementary to a sequence that encodes a HIO1005 polypeptide having at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95% or more sequence identity to the polypeptide sequence presented in SEQ ID NO:2.

In another embodiment a HIO1005 polypeptide of the invention comprises a polypeptide sequence with at least 50% or 60% identity to the HIO1005 polypeptide sequence of SEQ ID NO:2, and may have at least 70%, 80%, 85%, 90% or 95% or more sequence identity to the HIO1005 polypeptide sequence of SEQ ID NO:2, such as one or more protein kinase domains. In another embodiment, a HIO1005 polypeptide comprises a polypeptide sequence with at least 50%, 60%, 70%, 80%, 85%, 90% or 95% or more sequence identity to a functionally active fragment of the polypeptide presented in SEQ ID NO:2. In yet another embodiment, a HIO1005 polypeptide comprises a polypeptide sequence with at least 50%, 60%, 70%, 80%, or 90% identity to the polypeptide sequence of SEQ ID NO:2 over its entire length and comprises of one or more protein kinase domains.

In another aspect, a HIO1005 polynucleotide sequence is at least 50% to 60% identical over its entire length to the HIO1005 nucleic acid sequence presented as SEQ ID NO:1, or nucleic acid sequences that are complementary to such a HIO1005 sequence, and may comprise at least 70%, 80%, 85%, 90% or 95% or more sequence identity to the HIO1005 sequence presented as SEQ ID NO:1 or a functionally active fragment thereof, or complementary sequences.

As used herein, "percent (%) sequence identity" with respect to a specified subject sequence, or a specified portion thereof, is defined as the percentage of nucleotides or amino acids in the candidate derivative sequence identical with the nucleotides or amino acids in the subject sequence (or specified portion thereof), after aligning the sequences and introducing gaps, if necessary to achieve the maximum percent sequence identity, as generated by the program WU-BLAST-2.0a19 (Altschul *et al.*, J. Mol. Biol. (1997) 215:403-410) with search parameters set to default values. The HSP S and HSP S2 parameters are dynamic values and are established by the program itself depending upon the composition of the particular sequence and composition of the particular database against

which the sequence of interest is being searched. A "% identity value" is determined by the number of matching identical nucleotides or amino acids divided by the sequence length for which the percent identity is being reported. "Percent (%) amino acid sequence similarity" is determined by doing the same calculation as for determining % amino acid sequence identity, but including conservative amino acid substitutions in addition to identical amino acids in the computation. A conservative amino acid substitution is one in which an amino acid is substituted for another amino acid having similar properties such that the folding or activity of the protein is not significantly affected. Aromatic amino acids that can be substituted for each other are phenylalanine, tryptophan, and tyrosine; interchangeable hydrophobic amino acids are leucine, isoleucine, methionine, and valine; interchangeable polar amino acids are glutamine and asparagine; interchangeable basic amino acids are arginine, lysine and histidine; interchangeable acidic amino acids are aspartic acid and glutamic acid; and interchangeable small amino acids are alanine, serine, threonine, cysteine and glycine.

Derivative nucleic acid molecules of the subject nucleic acid molecules include sequences that selectively hybridize to the nucleic acid sequence of SEQ ID NO:1. The stringency of hybridization can be controlled by temperature, ionic strength, pH, and the presence of denaturing agents such as formamide during hybridization and washing. Conditions routinely used are well known (see, *e.g.*, Current Protocol in Molecular Biology, Vol. 1, Chap. 2.10, John Wiley & Sons, Publishers (1994); Sambrook *et al.*, Molecular Cloning, Cold Spring Harbor (1989)). In some embodiments, a nucleic acid molecule of the invention is capable of hybridizing to a nucleic acid molecule containing the nucleotide sequence of SEQ ID NO:1 under stringent hybridization conditions that are: prehybridization of filters containing nucleic acid for 8 hours to overnight at 65° C in a solution comprising 6X single strength citrate (SSC) (1X SSC is 0.15 M NaCl, 0.015 M Na citrate; pH 7.0), 5X Denhardt's solution, 0.05% sodium pyrophosphate and 100 µg/ml herring sperm DNA; hybridization for 18-20 hours at 65° C in a solution containing 6X SSC, 1X Denhardt's solution, 100 µg/ml yeast tRNA and 0.05% sodium pyrophosphate; and washing of filters at 65° C for 1 h in a solution containing 0.1X SSC and 0.1% SDS (sodium dodecyl sulfate). In other embodiments, moderately stringent hybridization conditions are used that are: pretreatment of filters containing nucleic acid for 6 h at 40° C in a solution containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.1% PVP, 0.1% Ficoll, 1% BSA, and 500 µg/ml denatured salmon sperm DNA; hybridization for 18-20 h at 40° C in a solution containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 µg/ml salmon sperm

DNA, and 10% (wt/vol) dextran sulfate; followed by washing twice for 1 hour at 55° C in a solution containing 2X SSC and 0.1% SDS. Alternatively, low stringency conditions can be used that comprise: incubation for 8 hours to overnight at 37° C in a solution comprising 20% formamide, 5 x SSC, 50 mM sodium phosphate (pH 7.6), 5X Denhardt's solution, 10%  
5 dextran sulfate, and 20 µg/ml denatured sheared salmon sperm DNA; hybridization in the same buffer for 18 to 20 hours; and washing of filters in 1 x SSC at about 37° C for 1 hour.

As a result of the degeneracy of the genetic code, a number of polynucleotide sequences encoding a HIO1005 polypeptide can be produced. For example, codons may be selected to increase the rate at which expression of the polypeptide occurs in a particular host  
10 species, in accordance with the optimum codon usage dictated by the particular host organism (see, e.g., Nakamura et al, 1999). Such sequence variants may be used in the methods of this invention.

The methods of the invention may use orthologs of the *Arabidopsis* HIO1005. Methods of identifying the orthologs in other plant species are known in the art. Normally,  
15 orthologs in different species retain the same function, due to presence of one or more protein motifs and/or 3-dimensional structures. In evolution, when a gene duplication event follows speciation, a single gene in one species, such as *Arabidopsis*, may correspond to multiple genes (paralogs) in another. As used herein, the term "orthologs" encompasses paralogs. When sequence data is available for a particular plant species, orthologs are  
20 generally identified by sequence homology analysis, such as BLAST analysis, usually using protein bait sequences. Sequences are assigned as a potential ortholog if the best hit sequence from the forward BLAST result retrieves the original query sequence in the reverse BLAST (Huynen MA and Bork P, Proc Natl Acad Sci (1998) 95:5849-5856; Huynen MA *et al.*, Genome Research (2000) 10:1204-1210).

Programs for multiple sequence alignment, such as CLUSTAL (Thompson JD et al, 1994, Nucleic Acids Res 22:4673-4680) may be used to highlight conserved regions and/or  
25 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  
30 tree with respect to all other sequences from these two species. Structural threading or other analysis of protein folding (e.g., using software by ProCeryon, Biosciences, Salzburg, Austria) may also identify potential orthologs. Nucleic acid hybridization methods may also be used to find orthologous genes and are preferred when sequence data are not available. Degenerate PCR and screening of cDNA or genomic DNA libraries are common methods  
35 for finding related gene sequences and are well known in the art (see, e.g., Sambrook, 1989;



Dieffenbach and Dveksler, 1995). For instance, methods for generating a cDNA library from the plant species of interest and probing the library with partially homologous gene probes are described in Sambrook *et al.* A highly conserved portion of the *Arabidopsis* HIO1005 coding sequence may be used as a probe. HIO1005 ortholog nucleic acids may hybridize to the nucleic acid of SEQ ID NO:1 under high, moderate, or low stringency conditions. After amplification or isolation of a segment of a putative ortholog, that segment may be cloned and sequenced by standard techniques and utilized as a probe to isolate a complete cDNA or genomic clone. Alternatively, it is possible to initiate an EST project to generate a database of sequence information for the plant species of interest. In another approach, antibodies that specifically bind known HIO1005 polypeptides are used for ortholog isolation (see, e.g., Harlow and Lane, 1988, 1999). Western blot analysis can determine that a HIO1005 ortholog (i.e., an orthologous protein) is present in a crude extract of a particular plant species. When reactivity is observed, the sequence encoding the candidate ortholog may be isolated by screening expression libraries representing the particular plant species. Expression libraries can be constructed in a variety of commercially available vectors, including lambda gt11, as described in Sambrook, *et al.*, 1989. Once the candidate ortholog(s) are identified by any of these means, candidate orthologous sequence are used as bait (the "query") for the reverse BLAST against sequences from *Arabidopsis* or other species in which HIO1005 nucleic acid and/or polypeptide sequences have been identified.

HIO1005 nucleic acids and polypeptides may be obtained using any available method. For instance, techniques for isolating cDNA or genomic DNA sequences of interest by screening DNA libraries or by using polymerase chain reaction (PCR), as previously described, are well known in the art. Alternatively, nucleic acid sequence may be synthesized. Any known method, such as site directed mutagenesis (Kunkel TA *et al.*, 1991), may be used to introduce desired changes into a cloned nucleic acid.

In general, the methods of the invention involve incorporating the desired form of the HIO1005 nucleic acid into a plant expression vector for transformation of in plant cells, and the HIO1005 polypeptide is expressed in the host plant.

An isolated HIO1005 nucleic acid molecule is other than in the form or setting in which it is found in nature and is identified and separated from least one contaminant nucleic acid molecule with which it is ordinarily associated in the natural source of the HIO1005 nucleic acid. However, an isolated HIO1005 nucleic acid molecule includes HIO1005 nucleic acid molecules contained in cells that ordinarily express HIO1005 where,

for example, the nucleic acid molecule is in a chromosomal location different from that of natural cells.

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

5 HIO1005 nucleic acids and polypeptides may be used in the generation of genetically modified plants having a modified oil content phenotype. As used herein, a "modified oil content phenotype" may refer to modified oil content in any part of the plant; the modified oil content is often observed in seeds. In a preferred embodiment, altered expression of the HIO1005 gene in a plant is used to generate plants with a high oil  
10 phenotype.

The methods described herein are generally applicable to all plants. Although activation tagging and gene identification is carried out in *Arabidopsis*, the HIO1005 gene (or an ortholog, variant or fragment thereof) may be expressed in any type of plant. In a preferred embodiment, the invention is directed to oil-producing plants, which produce and  
15 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  
20 (*Arachis hypogaea*). The invention may also be directed to fruit- and vegetable-bearing plants, grain-producing plants, nut-producing plants, rapid cycling *Brassica* species, alfalfa (*Medicago sativa*), tobacco (*Nicotiana*), turfgrass (*Poaceae* family), other forage crops, and wild species that may be a source of unique fatty acids.

The skilled artisan will recognize that a wide variety of transformation techniques  
25 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  
30 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  
35 passed onto successive plant generations. Depending upon the intended use, a heterologous

nucleic acid construct comprising an HIO1005 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 rapeseed (De Block *et al.*, 1989), sunflower (Everett *et al.*, 1987), and soybean (Christou *et al.*, 1989; Kline *et al.*, 1987).

Expression (including transcription and translation) of HIO1005 may be regulated with respect to the level of expression, the tissue type(s) where expression takes place and/or developmental stage of expression. A number of heterologous regulatory sequences (e.g., promoters and enhancers) are available for controlling the expression of a HIO1005 nucleic acid. These include constitutive, inducible and regulatable promoters, as well as promoters and enhancers that control expression in a tissue- or temporal-specific manner. Exemplary constitutive promoters include the raspberry E4 promoter (U.S. Patent Nos. 5,783,393 and 5,783,394), the 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 (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), the melon actin promoter (published PCT application WO0056863), 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

progeny of these plants that are not genetically modified. In one method, called "TILLING" (for targeting induced local lesions in genomes), mutations are induced in the seed of a plant of interest, for example, using EMS treatment. The resulting plants are grown and self-fertilized, and the progeny are used to prepare DNA samples. HIO1005-specific PCR is used to identify whether a mutated plant has a HIO1005 mutation. Plants having HIO1005 mutations may then be tested for altered oil content, or alternatively, plants may be tested for altered oil content, and then HIO1005-specific PCR is used to determine whether a plant having altered oil content has a mutated HIO1005 gene. TILLING can identify mutations that may alter the expression of specific genes or the activity of proteins encoded by these genes (see Colbert et al (2001) *Plant Physiol* 126:480-484; McCallum et al (2000) *Nature Biotechnology* 18:455-457).

In another method, a candidate gene/Quantitative Trait Locus (QTLs) approach can be used in a marker-assisted breeding program to identify alleles of or mutations in the HIO1005 gene or orthologs of HIO1005 that may confer altered oil content (see Bert et al., *Theor Appl Genet.* 2003 Jun;107(1):181-9; and Lionneton et al, *Genome.* 2002 Dec;45(6):1203-15). Thus, in a further aspect of the invention, a HIO1005 nucleic acid is used to identify whether a plant having altered oil content has a mutation in endogenous HIO1005 or has a particular allele that causes altered oil content.

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

## EXAMPLES

### EXAMPLE 1

#### Generation of Plants with a HIO1005 Phenotype by Transformation with an Activation

##### Tagging Construct

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

CaMV 35S enhancer element. Transgenic plants were selected at the T1 generation based on herbicide resistance.

T3 seed was analyzed by Near Infrared Spectroscopy (NIR) at the time of harvest. NIR infrared spectra were captured using a Bruker 22 N/F. Bruker Software was used to estimate total seed oil and total seed protein content using data from NIR analysis and reference methods according to the manufacturers instructions. Oil contents predicted by our calibration (ren oil 1473 1d + sline.q2, Predicts Hexane Extracted Oil), which followed the general method of AOCS Procedure AM1-92, Official Methods and Recommended Practices of the American Oil Chemists Society, 5th Ed., AOCS, Champaign, Ill., were compared for 38,090 individual ACTTAG lines. Subsequent to seed compositional analysis, the position of the ACTTAG element in the genome of each line was determined by inverse PCR and sequencing. 38,090 lines with recovered flanking sequences were considered in this analysis.

Since the 38,090 lines were planted and grown over a 12-month period, the seed oil content values were normalized to minimize the effect of environmental differences which may alter seed oil content. The average seed oil content and its standard deviation, for each day lines were planted, were calculated. The seed oil content was expressed as a "relative standard deviation distance" (SD distance) which was calculated by subtracting the average seed oil content for the planting day from seed oil content for each line and dividing the difference by the standard deviation for that day. This normalization allows comparison of seed oil content in seed from plants grown throughout the year.

Genes that cause a high seed oil phenotype when over-expressed were identified by evaluating all of the genes affected by ACTTAG elements in the 38,090 lines. This was accomplished by the following procedure; first, the genes likely to be activated by the ACTTAG element in each line were identified and the seed oil content of the line was assigned to these genes; second, the seed oil content when a particular gene is over-expressed was determined by averaging the individual seed oil values for each gene. Since 38,090 lines were evaluated and each element affects an average of 2.5 genes, each gene will have an average of 4 seed oil values. The genes with the highest average SD distance were determined to be those that cause a high seed oil phenotype when over-expressed.

Plants over-expressing At5g02290, HIO1005 have an oil content of 117% of the planting day average. The oil content and "relative standard deviation distance" of plants over-expressing At5g02290, HIO1005 is shown in the following Table 1.

**Table 1.**

Tair	relative standard deviation distance	plant count	Description	Line ID	Planting date	n (# of ACTTAG lines planted for the date with NIR measurement)	Seed Oil content (%)	Average oil content of the planting date (%)	Standard deviation of the oil content for the planting date
At5g02290	2.764666	1	protein kinase, putative	W000177287	7/31/02	1102	38.543	32.968	2.016

EXAMPLE 25 Characterization of the T-DNA Insertion in Plants Exhibiting the Altered Oil Content Phenotype.

We performed standard molecular analyses, essentially as described in patent application PCT WO0183697, to determine the site of the T-DNA insertion associated with the altered oil content phenotype. Briefly, genomic DNA was extracted from plants exhibiting the altered oil content phenotype. PCR, using primers specific to the pSKI015 vector, confirmed the presence of the 35S enhancer in plants from the HIO1005 oil line, and Southern blot analysis verified the genomic integration of the ACTTAG T-DNA and showed the presence of the T-DNA insertions in each of the transgenic lines.

Inverse PCR was used to recover genomic DNA flanking the T-DNA insertion, which was then subjected to sequence analysis using a basic BLASTN search and/or a search of the *Arabidopsis* Information Resource (TAIR) database (available at the arabidopsis.org website).

EXAMPLE 320 Recapitulation of HIO1005 phenotype

To test whether over-expression of At5g02290 causes a high seed oil phenotype, oil content in seeds from transgenic plants over-expressing this gene was compared with oil content in seeds from non-transgenic control plants. To do this, At5g02290 was cloned into a plant transformation vector behind the seed specific CsVMV promoter and transformed into *Arabidopsis* plants using the floral dip method. The plant transformation vector contains the nptII gene driven by the RE4 promoter, to provide resistance to kanamycin, and serve as a selectable marker. Seed from the transformed plants were plated on agar medium containing kanamycin. 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 7 days and then transplanted to soil. Twenty-two transgenic seedlings and 10 non-transgenic control plants were transplanted to random positions in the same 32 cell flat. 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 percent oil in the seed harvested from each plant as determined by NIR spectroscopy is presented in Table 3. The Relative Oil value is determined by dividing the predicted oil value by the average oil value in control seed (i.e. seed from plants without the transgene).

The effect of over-expression of At5g02290 on seed oil has been tested in two experiments. In both experiments, the plants over-expressing At5g02290 had higher seed oil content than the control plants grown in the same flat. Across the experiments, the average seed oil content of plants over-expressing At5g02290 was 3.9% greater than the untransformed controls. The seed oil content in plants over-expressing At5g02290 was significantly greater than non-transgenic control plants (two-way ANOVA;  $P = 0.0297$ ).

**Table 2**

Experiment	Plant	Transgene	Percent Oil	Relative Oil
1	DX07125001	CsVMV:At5g02290	25.28	96.36
1	DX07125002	CsVMV:At5g02290	27.75	105.76
1	DX07125003	CsVMV:At5g02290	27.65	105.41
1	DX07125004	CsVMV:At5g02290	29.35	111.88
1	DX07125005	CsVMV:At5g02290	26.19	99.83
1	DX07125006	CsVMV:At5g02290	30.18	115.03
1	DX07125007	CsVMV:At5g02290	24.03	91.59
1	DX07125008	CsVMV:At5g02290	25.46	97.02
1	DX07125010	CsVMV:At5g02290	25.28	96.34
1	DX07125011	CsVMV:At5g02290	28.39	108.21
1	DX07125012	CsVMV:At5g02290	27.26	103.92
1	DX07125013	CsVMV:At5g02290	25.44	96.98
1	DX07125014	CsVMV:At5g02290	27.07	103.18
1	DX07125015	CsVMV:At5g02290	32.84	125.17
1	DX07125016	CsVMV:At5g02290	30.62	116.71
1	DX07125017	CsVMV:At5g02290	25.98	99.03
1	DX07125018	CsVMV:At5g02290	27.99	106.69
1	DX07125019	CsVMV:At5g02290	26.33	100.37
1	DX07125020	CsVMV:At5g02290	26.35	100.44
1	DX07125021	CsVMV:At5g02290	27.21	103.7
1	DX07125022	CsVMV:At5g02290	25.46	97.04
1	DX07143001	None	28.38	108.16
1	DX07143002	None	26.94	102.68
1	DX07143003	None	25.27	96.33

1	DX07143004	None	25.14	95.81
1	DX07143005	None	26.6	101.39
1	DX07143006	None	27.45	104.63
1	DX07143007	None	24.31	92.66
1	DX07143008	None	27.15	103.48
1	DX07143009	None	25.13	95.8
1	DX07143010	None	25.99	99.07
2	DX07170001	CsVMV:At5g02290	33.73	111.06
2	DX07170002	CsVMV:At5g02290	32.28	106.3
2	DX07170003	CsVMV:At5g02290	30.78	101.35
2	DX07170004	CsVMV:At5g02290	33.9	111.63
2	DX07170005	CsVMV:At5g02290	32.84	108.15
2	DX07170006	CsVMV:At5g02290	29.64	97.59
2	DX07170007	CsVMV:At5g02290	32.29	106.33
2	DX07170008	CsVMV:At5g02290	31.41	103.43
2	DX07170009	CsVMV:At5g02290	30.56	100.62
2	DX07170010	CsVMV:At5g02290	35.85	118.04
2	DX07170011	CsVMV:At5g02290	28.19	92.82
2	DX07170012	CsVMV:At5g02290	31.78	104.65
2	DX07170013	CsVMV:At5g02290	33.64	110.78
2	DX07170014	CsVMV:At5g02290	28.66	94.36
2	DX07170015	CsVMV:At5g02290	33.56	110.51
2	DX07170016	CsVMV:At5g02290	31.2	102.74
2	DX07170017	CsVMV:At5g02290	30.75	101.24
2	DX07170019	CsVMV:At5g02290	30.86	101.61
2	DX07170020	CsVMV:At5g02290	30.42	100.18
2	DX07170021	CsVMV:At5g02290	29.68	97.74
2	DX07188001	None	30.1	99.12
2	DX07188002	None	29.25	96.3
2	DX07188003	None	30.44	100.22
2	DX07188004	None	31.06	102.27
2	DX07188005	None	30.1	99.12
2	DX07188006	None	31.03	102.17
2	DX07188007	None	29.68	97.72
2	DX07188008	None	29.12	95.89
2	DX07188010	None	32.55	107.19

**EXAMPLE 4**Analysis of *Arabidopsis* HIO1005 Sequence

- 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), PSORT (Nakai K, and Horton P, 1999, Trends Biochem Sci 24:34-6), and/or CLUSTAL (Thompson JD *et al.*, 1994, Nucleic Acids Res 22:4673-4680).



**TBLASTN against ESTs:**

The candidate gene At5g02290 is supported by the full-length cDNA gi|42458510, gi|42458347, gi|42458159, gi|42457560 gi|42457287. Note that the 2 protein sequence entries for At5g02290 (>gi|15241749 and >gi|30679590) are 100% identical and therefore  
5 give exactly the same computational analysis results described below.

There are many ESTs from diverse plant species showing similarity to At5g02290. Where possible, ESTs contigs of each species were made. The top hit for each of the following species are listed below and included in the orthologue Table 2: Triticum aestivum, Glycine max, Mentha x piperita, Populus tremula, Oryza sativa, Lycopersicon  
10 esculentum, Solanum tuberosum, and Beta vulgaris.

**1. Sugar Beet ESTs with the following GenBank IDs:**

gi|34892113

**15 2. Potato ESTs with the following GenBank IDs:**

gi|9249744

gi|10446230

gi|12588854

gi|14642480

**20** gi|17070644

gi|17070774

gi|17070788

gi|17075314

gi|19821257

**25** gi|21922218

gi|21375853

**3. Tomato ESTs with the following GenBank IDs:**

gi|4381266

**30** gi|5602522

gi|5897320

gi|4381266

gi|5602522

gi|5897320

**35** gi|5897320

gi|14688536

gi|14688536

**40 4. Rice ESTs with the following GenBank IDs:**

gi|2317369

gi|2317406

gi|5443728

**45 5. Poplar ESTs with the following GenBank IDs:**

gi|23985034

gi|24060873

gi|28609471

gi|27421455

**6. Mint ESTs with the following GenBank IDs:**

gi|7244808

5

**7. Soybean ESTs with the following GenBank IDs:**

gi|7478486

gi|7638957

gi|7925466

10

gi|13790463

gi|15812567

gi|15815577

gi|16344047

gi|16344695

15

gi|19270028

gi|22525920

gi|22525985

gi|27446266

20

**8. Corn ESTs with the following GenBank IDs:**

gi|5915602

gi|6194459

gi|6826956

gi|6828040

25

gi|21210829

gi|18173600

gi|18650074

gi|22472730

gi|22473455

30

gi|22545925

gi|23199478

gi|26558828

gi|26558898

35

**9. Cotton ESTs with the following GenBank IDs:**

gi|5050902

gi|5050715

gi|5046977

40

**10. Wheat ESTs with the following GenBank IDs:**

gi|11543439

gi|20102206

45

gi|20106492

gi|20113019

gi|17145892

gi|21483048

gi|22212150

50

gi|22302945

gi|32556607

**BLASTP against amino acids:**

The protein At5g02290 has homology to protein kinases from plants, fungi and animals. Detailed sequence analysis of protein kinases from plant showed that At5g02290 is a member of the Family 1.2.2 of plant protein kinase (Receptor Like Cytoplasmic serine/threonine Kinase VII) of which there are 51 members (PPC family 1.2.2). These proteins are likely to perform similar functions in plants. The top 10 BLAST results for At5g02290 are listed below and are included in the Orthologue Table 2 below.

10           **1. At5g02290 itself**

>gi|15241749|ref|NP\_195849.1| protein kinase, putative [Arabidopsis thaliana]  
 >gi|30679590|ref|NP\_850755.1| protein kinase, putative [Arabidopsis thaliana]  
 >gi|21431784|sp|P43293|NAK\_ARATH Probable serine/threonine-protein kinase NAK  
 >gi|11376624|pir||T48250 serine/threonine-specific protein kinase NAK (EC 2.7.1.-) -  
 15 Arabidopsis thaliana >gi|7406425|emb|CAB85534.1| serine/threonine-specific protein  
 kinase NAK [Arabidopsis thaliana] >gi|21555255|gb|AAM63816.1| serine/threonine-  
 specific protein kinase NAK [Arabidopsis thaliana]

Length = 389

Score = 2017 (715.1 bits), Expect = 9.9e-208, P = 9.9e-208

20 Identities = 389/389 (100%), Positives = 389/389 (100%)

The following sequence is another redundant entry of At5g02290. However, it differs from the sequence listed above by a few nucleotides. This is likely to be the result of sequencing errors or of single nucleotide polymorphisms with little or no effect on activity.

>gi|166809|gb|AAA18853.1| protein kinase

25 Length = 389

Score = 2008 (711.9 bits), Expect = 8.9e-207, P = 8.9e-207

Identities = 388/389 (99%), Positives = 388/389 (99%)

**2. putative serine/threonine-specific protein kinase from rice**

>gi|34906046|ref|NP\_914370.1| putative serine/threonine-specific protein kinase [Oryza  
 30 sativa (japonica cultivar-group)] >gi|15290002|dbj|BAB63697.1| putative serine/threonine-  
 specific protein kinase [Oryza sativa (japonica cultivar-group)]  
 >gi|20160935|dbj|BAB89871.1| putative serine/threonine-specific protein kinase [Oryza  
 sativa (japonica cultivar-group)]

Length = 467

35 Score = 1331 (473.6 bits), Expect = 4.9e-135, P = 4.9e-135

Identities = 252/375 (67%), Positives = 307/375 (81%)

**3. At1g07570 from Arabidopsis**

>gi|15222437|ref|NP\_172237.1| protein kinase (APK1a) [Arabidopsis thaliana]

>gi|42571375|ref|NP\_973778.1| protein kinase (APK1a) [Arabidopsis thaliana]

>gi|1168470|sp|Q06548|APKA\_ARATH Protein kinase APK1A >gi|282877|pir||S28615

5 serine/threonine/tyrosine-specific protein kinase APK1 (EC 2.7.1.-) [validated] -

Arabidopsis thaliana >gi|217829|dbj|BAA02092.1| protein tyrosine-serine-threonine kinase

[Arabidopsis thaliana] >gi|28393320|gb|AAO42086.1| putative protein kinase APK1A

[Arabidopsis thaliana] >gi|28827602|gb|AAO50645.1| putative protein kinase APK1A

[Arabidopsis thaliana]

10 Length = 410

Score = 1330 (473.2 bits), Expect = 6.3e-135, P = 6.3e-135

Identities = 253/359 (70%), Positives = 303/359 (84%)

The following sequence is another redundant entry of At1g07570. However, it differs from the sequence listed above by a few nucleotides. This is likely to be the result of

15 sequencing errors or of single nucleotide polymorphisms with little or no effect on activity.

>gi|8778537|gb|AAF79545.1| F22G5.5 [Arabidopsis thaliana]

Length = 464

Score = 993 (354.6 bits), Expect = 1.9e-133, Sum P(2) = 1.9e-133

Identities = 192/277 (69%), Positives = 229/277 (82%)

20 **4. At2g28930 from Arabidopsis**

>gi|42569425|ref|NP\_180459.2| protein kinase (APK1b) [Arabidopsis thaliana]

Length = 423

Score = 1314 (467.6 bits), Expect = 3.1e-133, P = 3.1e-133

Identities = 252/373 (67%), Positives = 301/373 (80%)

25 The following sequences are other redundant entries of At2g28930. However, they differ from the sequence listed above by a few nucleotides. This is likely to be the result of sequencing errors or of single nucleotide polymorphisms with little or no effect on activity.

>gi|12644274|sp|P46573|APKB\_ARATH Protein kinase APK1B >gi|7434286|pir||T02725

probable serine/threonine/tyrosine-specific protein kinase (EC 2.7.1.-) T9I4.1 - Arabidopsis

30 thaliana >gi|3461835|gb|AAC33221.1| putative protein kinase [Arabidopsis thaliana]

>gi|20197437|gb|AAM15075.1| putative protein kinase [Arabidopsis thaliana]

Length = 412

Score = 1327 (472.2 bits), Expect = 1.3e-134, P = 1.3e-134

Identities = 258/388 (66%), Positives = 309/388 (79%)

**5. putative protein kinase from rice**

>gi|14719339|gb|AAK73157.1| putative protein kinase [Oryza sativa]

Length = 395

Score = 1310 (466.2 bits), Expect = 8.2e-133, P = 8.2e-133

5 Identities = 249/370 (67%), Positives = 302/370 (81%)

**6. putative protein kinase from rice**

>gi|34897424|ref|NP\_910058.1| putative protein kinase [Oryza sativa (japonica cultivar-group)] >gi|27545044|gb|AAO18450.1| putative protein kinase [Oryza sativa (japonica cultivar-group)]

10 Length = 416

Score = 1284 (457.0 bits), Expect = 4.7e-130, P = 4.7e-130

Identities = 254/402 (63%), Positives = 321/402 (79%)

**7. At2g39660 from Arabidopsis**

>gi|15225520|ref|NP\_181496.1| protein kinase, putative [Arabidopsis thaliana]

15 >gi|7434287|pir||T00574 probable protein kinase [imported] - Arabidopsis thaliana

>gi|2795805|gb|AAB97121.1| putative protein kinase [Arabidopsis thaliana]

>gi|13272431|gb|AAK17154.1| putative protein kinase [Arabidopsis thaliana]

>gi|17064834|gb|AAL32571.1| putative protein kinase [Arabidopsis thaliana]

>gi|18086424|gb|AAL57667.1| At2g39660/F12L6.32 [Arabidopsis thaliana]

20 >gi|20197111|gb|AAM14921.1| putative protein kinase [Arabidopsis thaliana]

>gi|20259860|gb|AAM13277.1| putative protein kinase [Arabidopsis thaliana]

>gi|20334794|gb|AAM16258.1| At2g39660/F12L6.32 [Arabidopsis thaliana]

Length = 395

Score = 1272 (452.8 bits), Expect = 8.8e-129, P = 8.8e-129

25 Identities = 241/368 (65%), Positives = 300/368 (81%)

**8. protein kinase from canola (Brassica)**

>gi|45259527|dbj|BAD12263.1| protein kinase [Brassica rapa]

Length = 404

Score = 1252 (445.8 bits), Expect = 1.2e-126, P = 1.2e-126

30 Identities = 244/398 (61%), Positives = 301/398 (75%)

**9. At3g55450 from Arabidopsis**

>gi|30694253|ref|NP\_191105.2| protein kinase, putative [Arabidopsis thaliana]

>gi|23306362|gb|AAN17408.1| serine/threonine-specific protein kinase -like [Arabidopsis thaliana]

35 Length = 389

Score = 1179 (420.1 bits), Expect = 6.3e-119, P = 6.3e-119

Identities = 226/354 (63%), Positives = 283/354 (79%)

**10. putative protein tyrosine-serine-threonine kinase from rice**

>gi|19071648|gb|AAL84315.1| putative protein tyrosine-serine-threonine kinase [Oryza  
5 sativa (japonica cultivar-group)]

Length = 422

Score = 1178 (419.7 bits), Expect = 8.0e-119, P = 8.0e-119

Identities = 235/394 (59%), Positives = 301/394 (76%)

10 **Table 3.**

Ortholog Gene Name	Species	GI #	% ID to HIO1005	Score(s) (BLAST, Clustal, etc.)
One EST contig from potato	<b>Solanum tuberosum</b>	gi 9249744 gi 10446230 gi 12588854 gi 14642480 gi 17070644 gi 17070774 gi 17070788 gi 17075314 gi 19821257 gi 21922218 gi 21375853 consensus: SEQ ID NO:5	Length: 2205 Identities: 0.673 Positives: 0.822 Frames: 2	TBLASTN Score: 1404 Probability: 1.700000e-144
One EST contig from sugar beet	<b>Beta vulgaris</b>	gi 34892113	Length: 674 Identities: 0.664 Positives: 0.818 Frames: 2	TBLASTN Score: 737 Probability: 1.100000e-73
One EST contig from soybean	<b>Glycine max</b>	gi 7478486 gi 7638957 gi 7925466 gi 13790463 gi 15812567 gi 15815577 gi 16344047 gi 16344695 gi 19270028 gi 22525920 gi 22525985 gi 27446266 consensus: SEQ ID NO:9	Length: 1631 Identities: 0.669 Positives: 0.817 Frames: 2	TBLASTN Score: 1378 Probability: 2.900000e-141

Ortholog Gene Name	Species	GI #	% ID to HIO1005	Score(s) (BLAST, Clustal, etc.)
One EST from wheat	<b>Triticum aestivum</b>	gi 11543439 gi 20102206 gi 20106492 gi 20113019 gi 17145892 gi 21483048 gi 22212150 gi 22302945 gi 32556607 consensus: SEQ ID NO:12	Length: 1921  Identities: 0.613 Positives: 0.768 Frames: 3	TBLASTN Score: 1178 Probability: 1.600000e-122
One EST contig from tomato	<b>Lycopersicon esculentum</b>	gi 4381266 gi 5602522 gi 5897320 gi 4381266 gi 5602522 gi 5897320 gi 5897320 gi 14688536 gi 14688536 consensus: SEQ ID NO:6	Length: 1601  Identities: 0.608 Positives: 0.791  Frames: 2	TBLASTN Score: 1270 Probability: 6.300000e-130
One EST contig from rice	<b>Oryza sativa</b>	gi 2317369 gi 2317406 gi 5443728 consensus: SEQ ID NO:7	Length: 965  Identities: 0.630 Positives: 0.768  Frames: 3	TBLASTN Score: 781 Probability: 9.700000e-90
One EST contig from corn	<b>Zea mays</b>	gi 5915602 gi 6194459 gi 6826956 gi 6828040 gi 21210829 gi 18173600 gi 18650074 gi 22472730 gi 22473455 gi 22545925 gi 23199478 gi 26558828 gi 26558898 consensus: SEQ ID NO:10	Length: 1620  Identities: 0.676 Positives: 0.838 Frames: -3	TBLASTN Score: 1258 Probability: 9.700000e-129
One EST contig from cotton	<b>Gossypium hirsutum</b>	gi 5050902 gi 5050715 gi 5046977 consensus: SEQ ID NO:11	Length: 632  Identities: 0.597 Positives: 0.761  Frames: 3	TBLASTN Score: 539 Probability: 6.800000e-53
One EST contig from poplar	<b>Populus tremula</b>	gi 23985034 gi 24060873 gi 28609471 gi 27421455 consensus: SEQ ID NO:8	Length: 1240  Identities: 0.684 Positives: 0.829  Frames: 3	TBLASTN Score: 958 Probability: 5.000000e-97

Ortholog Gene Name	Species	GI #	% ID to HIO1005	Score(s) (BLAST, Clustal, etc.)
One EST contig from mint	<b>Mentha piperita</b>	gi 7244808	Length: 646 Identities: 0.470 Positives: 0.722 Frames: 1	TBLASTN Score: 262 Probability: 2.600000e-24

**Closest plant homologs:**

At5g02290	<b>Arabidopsis thaliana</b>	gi 15241749 gi 30679590 gi 21431784 gi 11376624 gi 7406425 gi 21555255	Identities = 389/389 (100%), Positives = 389/389 (100%) Frames: N	<i>BLASTP</i> Score: 2017 P = 9.9e-208
Serine/threonine-specific protein kinase from rice	<b>Oryza sativa (japonica cultivar-group)</b>	gi 34906046 gi 15290002 gi 20160935	Identities = 252/375 (67%), Positives = 307/375 (81%) Frames: N	<i>BLASTP</i> Score: 1331 P = 4.9e-135
At1g07570	<b>Arabidopsis thaliana</b>	gi 15222437 gi 42571375 gi 1168470 gi 282877 gi 217829 gi 28393320 gi 28827602	Identities = 253/359 (70%), Positives = 303/359 (84%) Frames: N	<i>BLASTP</i> Score: 1330 P = 6.3e-135
At2g28930	<b>Arabidopsis thaliana</b>	gi 42569425	Identities = 252/373 (67%), Positives = 301/373 (80%) Frames: N	<i>BLASTP</i> Score: 1314 P = 3.1e-133
Protein kinase from rice	<b>Oryza sativa</b>	gi 14719339	Identities = 249/370 (67%), Positives = 302/370 (81%) Frames: N	<i>BLASTP</i> Score: 1310 P = 8.2e-133
Protein kinase from rice	<b>Oryza sativa (japonica cultivar-group)</b>	gi 34897424 gi 27545044	Identities = 254/402 (63%), Positives = 321/402 (79%) Frames: N	<i>BLASTP</i> Score: 1284 P = 4.7e-130
Protein kinase from rice	<b>Arabidopsis thaliana</b>	gi 15225520 gi 7434287 gi 2795805 gi 13272431 gi 17064834 gi 18086424 gi 20197111 gi 20259860 gi 20334794	Identities = 241/368 (65%), Positives = 300/368 (81%) Frames: N	<i>BLASTP</i> Score: 1272 P = 8.8e-129
Protein kinase from canola	<b>Brassica rapa</b>	gi 45259527	Identities = 244/398 (61%), Positives = 301/398 (75%) Frames: N	<i>BLASTP</i> Score: 1252 P = 1.2e-126



At3g55450	<b>Arabidopsis thaliana</b>	gij30694253 gij23306362	Identities = 226/354 (63%), Positives = 283/354 (79%) Frames: N	<i>BLASTP</i> Score: 1179 P = 6.3e-119
Protein tyrosine-serine-threonine kinase	<b>Oryza sativa (japonica cultivar-group)</b>	gij19071648	Identities = 235/394 (59%), Positives = 301/394 (76%) Frames: N	<i>BLASTP</i> Score: 1178 P = 8.0e-119

At5g02290 may be a chloroplastic protein (predicted by TargetP). Consistent with this result, At5g02290 lacks signal peptide and lacks trans-membrane domain as predicted by SignalP and TMHMM, respectively.

- 5 Pfam analysis showed that At5g02290 is a member of the serine/threonine protein kinase family (COG0515, SM00220, PF00069).

Model	Domain	seq-f*	seq-t	hmm-f	hmm-t	score	E-value
10 COG0515	1/1	68	388 ..	1	536 []	88.3	3.1e-23
SM00220	1/1	68	348 ..	1	231 []	140.7	5.2e-39
PF00069	1/1	68	350 ..	1	278 []	182.0	2e-51

- 15 \*Seq-f refers to “sequence-from” and seq-t refers to “sequence-to.” The two periods following the seq-t number indicate that the matching region was within the sequence and did not extend to either end. The two brackets indicate that the match spanned the entire length of the profile HMM. hmm-f and hmm-t refer to the beginning and ending coordinates of the matching portion of the profile HMM.

- 20 Eukaryotic protein kinases are enzymes that belong to a very extensive family of proteins which share a conserved catalytic core common with both serine/threonine and tyrosine protein kinases. There are a number of conserved regions in the catalytic domain of protein kinases. In the N-terminal extremity of the catalytic domain there is a glycine-rich stretch of residues in the vicinity of a lysine residue, which has been shown to be involved
- 25 in ATP binding. In the central part of the catalytic domain there is a conserved aspartic acid residue which is important for the catalytic activity of the enzyme. While the SMART signature, SM00220, identifies the catalytic domain, the PROSITE signature, PS00108, identifies the active site.

- 30 Detailed sequence analysis of protein kinases from plant showed that At5g02290 is a member of the Family 1.2.2 of plant protein kinase (Receptor Like Cytoplasmic serine/threonine Kinase VII) of which there are 51 members (PPC family 1.2.2). Note that the classification of protein kinase by plantsP is based on sequence comparison using BLAST. The classification system does not take into account of the cellular localization of proteins. Therefore, although Family 1.2.2 of plant protein kinase is called “Receptor Like

Cytoplasmic Kinase VII", no implication is made regarding the cellular localization (e.g. cytoplasmic, nuclear, trans-membrane) of its members.

In conclusion, the chloroplastic protein kinase At5g02290 may regulate activities of other proteins that may have an impact on the fatty acid synthesis.

5

#### EXAMPLE 5

Transformed explants of rapeseed, soy, corn, sunflower, cotton, cocoa, safflower, oil palm, coconut palm, flax, castor and peanut are obtained through *Agrobacterium tumefaciens*-mediated transformation or microparticle bombardment. Plants are regenerated from transformed tissue. The greenhouse grown plants are then analyzed for the gene of interest expression levels as well as oil levels.

10

#### EXAMPLE 6

This example provides analytical procedures to determine oil and protein content, mass differences, amino acid composition, free amino acid levels, and micronutrient content of transgenic maize plants.

15

Oil levels (on a mass basis and as a percent of tissue weight) of first generation single corn kernels and dissected germ and endosperm are determined by low-resolution <sup>1</sup>H nuclear magnetic resonance (NMR) (Tiwari *et al.*, *JAOCS*, 51:104-109 (1974); or Rubel, *JAOCS*, 71:1057-1062 (1994)), whereby NMR relaxation times of single kernel samples are measured, and oil levels are calculated based on regression analysis using a standard curve generated from analysis of corn kernels with varying oil levels as determined gravimetrically following accelerated solvent extraction. One-way analysis of variance and the Student's T-test (JMP, version 4.04, SAS Institute Inc., Cary, NC, USA) are performed to identify significant differences between transgenic and non-transgenic kernels as determined by transgene-specific PCR.

20

25

Oil levels and protein levels in second generation seed are determined by NIT spectroscopy, whereby NIT spectra of pooled seed samples harvested from individual plants are measured, and oil and protein levels are calculated based on regression analysis using a standard curve generated from analysis of corn kernels with varying oil or protein levels, as determined gravimetrically following accelerated solvent extraction or elemental (%N) analysis, respectively. One-way analysis of variance and the Student's T-test are performed to identify significant differences in oil (% kernel weight) and protein (% kernel weight) between seed from marker positive and marker negative plants.

30

The levels of free amino acids are analyzed from each of the transgenic events using the following procedure. Seeds from each of the transgenic plants are crushed individually into a fine powder and approximately 50 mg of the resulting powder is transferred to a pre-weighed centrifuge tube. The exact sample weight is recorded and 1.0 ml of 5% trichloroacetic acid is added to each sample tube. The samples are mixed at room temperature by vortex and then centrifuged for 15 minutes at 14,000 rpm on an Eppendorf microcentrifuge (Model 5415C, Brinkmann Instrument, Westbury, NY). An aliquot of the supernatant is removed and analyzed by HPLC (Agilent 1100) using the procedure set forth in Agilent Technical Publication "Amino Acid Analysis Using the Zorbax Eclipse-AAA Columns and the Agilent 1100 HPLC," March 17, 2000.

Quantitative determination of total amino acids from corn is performed by the following method. Kernels are ground and approximately 60 mg of the resulting meal is acid-hydrolyzed using 6 N HCl under reflux at 100°C for 24 hrs. Samples are dried and reconstituted in 0.1 N HCl followed by precolumn derivatization with  $\alpha$ -phthalaldehyde (OPA0 for HPLC analysis. The amino acids are separated by a reverse-phase Zorbax Eclipse XDB-C18 HPLC column on an Agilent 1100 HPLC (Agilent, Palo Alto, CA). The amino acids are detected by fluorescence. Cysteine, proline, asparagine, glutamine, and tryptophan are not included in this amino acid screen (Henderson *et al.*, "Rapid, Accurate, Sensitive and Reproducible HPLC Analysis of Amino acids, Amino Acid Analysis Using Zorbax Eclipse-AAA Columns and the Agilent 1100 HPLC," Agilent Publication (2000); *see*, also, "Measurement of Acid-Stable Amino Acids," AACC Method 07-01 (American Association of Cereal Chemists, Approved Methods, 9th edition (LCCC# 95-75308)). Total tryptophan is measured in corn kernels using an alkaline hydrolysis method as described (Approved Methods of the American Association of Cereal Chemists –10<sup>th</sup> edition, AACC ed, (2000) 07-20 Measurement of Tryptophan – Alakline Hydrolysis).

Tocopherol and tocotrienol levels in seeds are assayed by methods well-known in the art. Briefly, 10 mg of seed tissue are added to 1 g of microbeads (Biospec Product Inc, Barlesville, OK) in a sterile microfuge tube to which 500  $\mu$ l 1% pyrogallol (Sigma Chemical Co., St. Louis, MO)/ethanol have been added. The mixture is shaken for 3 minutes in a mini Beadbeater (Biospec) on "fast" speed, then filtered through a 0.2  $\mu$ m filter into an autosampler tube. The filtered extracts are analyzed by HPLC using a Zorbax silica HPLC column (4.6 mm $\times$ 250 mm) with a fluorescent detection, an excitation at 290 nm, an emission at 336 nm, and bandpass and slits. Solvent composition and running conditions are as listed below with solvent A as hexane and solvent B as methyl-t-butyl ether. The injection volume is 20  $\mu$ l, the flow rate is 1.5 ml/minute and the run time is 12 minutes at

40°C. The solvent gradient is 90% solvent A, 10% solvent B for 10 minutes; 25% solvent A, 75% solvent B for 11 minutes; and 90% solvent A, 10% solvent B for 12 minutes.

Tocopherol standards in 1% pyrogallol/ethanol are run for comparison ( $\alpha$ -tocopherol,  $\gamma$ -tocopherol,  $\beta$ -tocopherol,  $\delta$ -tocopherol, and tocopherol (tocol)). Standard curves for alpha,  
5 beta, delta, and gamma tocopherol are calculated using Chemstation software (Hewlett Packard). Tocotrienol standards in 1% pyrogallol/ethanol are run for comparison ( $\alpha$ -tocotrienol,  $\gamma$ -tocotrienol,  $\beta$ -tocotrienol,  $\delta$ -tocotrienol). Standard curves for  $\alpha$ -,  $\beta$ -,  $\delta$ -, and  $\gamma$ -tocotrienol are calculated using Chemstation software (Hewlett Packard).

Carotenoid levels within transgenic corn kernels are determined by a standard  
10 protocol (Craft, *Meth. Enzymol.*, 213:185-205 (1992)). Plastoquinols and phylloquinones are determined by standard protocols (Threlfall *et al.*, *Methods in Enzymology*, XVIII, part C, 369-396 (1971); and Ramadan *et al.*, *Eur. Food Res. Technol.*, 214(6):521-527 (2002)).

**REFERENCES**

- Altschul, S.F. *et al.*, J. Mol. Biol. 215:403-410, 1990.
- Altschul, S.F. *et al.*, Nucleic Acids Res. 25:3389-3402, 1997.
- 5 Ausubel FM *et al.*, Current Protocols in Molecular Biology, John Wiley & Sons, New York, N.Y., 1993.
- Baldwin D *et al.*, Cur Opin Plant Biol. 2(2):96-103, 1999.
- Bateman *et al.*, 1999, Nucleic Acids Res 27:260-262 (website at pfam.wustl.edu).
- Baulcombe D, Arch Virol Suppl 15:189-201, 1999.
- 10 Cannon *et al.*, Plant Molec. Biol. (1990) 15:39-47.
- Ch'ng *et al.*, Proc. Natl. Acad. Sci. USA (1989) 86:10006-10010
- Christensen S *et al.*, 9<sup>th</sup> International Conference on *Arabidopsis* Research. Univ. of Wisconsin-Madison, June 24-28, 1998. Abstract 165.
- Christou *et al.*, Proc. Natl. Acad. Sci USA (1989) 86:7500-7504.
- 15 De Block *et al.*, Plant Physiol. (1989) 91:694-701.
- Dieffenbach C and Dveksler G (Eds.) PCR Primer: A Laboratory Manual, Cold Spring Harbor Laboratory Press, NY, 1989.
- Everett *et al.*, Bio/Technology (1987) 5:1201
- Feldmann *et al.*, Science 243: 1351-1354, 1989.
- 20 Focks N and Benning C, Plant Physiol 118:91-101, 1998.
- Fridborg I *et al.*, Plant Cell 11: 1019-1032, 1999.
- Harlow E and Lane D, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1988, New York.
- Harlow E and Lane D, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1999, New York
- 25 Hayashi H *et al.*, Science 258: 1350-1353, 1992.
- Jako *et al.*, Plant Physiology 126(2):861-74, 2001.
- James DW and Dooner HK (1990) Theor Appl Genet 80, 241-245.
- Jones JD *et al.*, Transgenic Res 1:285-297 1992.
- 30 Kardailsky I *et al.*, Science 286: 1962-1965, 1999.
- Katavic V *et al.*, Plant Physiology 108(1):399-409, 1995.
- Kline *et al.*, Nature (1987) 327:70.
- Kunkel TA *et al.*, Methods Enzymol. 204:125-39, 1991.
- Lemieux B., *et al.*, 1990, Theor Appl Genet 80, 234-240.
- 35 Nakamura Y *et al.*, 1999, Nucleic Acids Res 27:292.

- Napoli, *et al.*, *Plant Cell* 2:279-289, 1990.
- Okuley *et al.*, *Plant Cell* 6(1):147-158, 1994.
- Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual (Second Edition)*, Cold Spring Harbor Press, Plainview, N.Y., 1989.
- 5 Schaffer R, *et al.*, *Cell* 93: 1219-1229, 1998.
- Sheehy *et al.*, *Proc. Natl. Acad. Sci. USA* (1988) 85:8805-8809.
- Smith, *et al.*, *Nature* 334:724-726, 1988.
- Smith *et al.*, *Mol. Gen. Genetics* (1990) 224:477-481.
- Thompson JD *et al.*, *Nucleic Acids Res* 22:4673-4680, 1994.
- 10 van der Krol *et al.*, *Biotechniques* (1988) 6:958-976.
- van der Krol *et al.*, *The Plant Cell* (1990) 2:291-299.
- Van Haaren MJJ *et al.*, *Plant Mol Bio* 21:625-640, 1993.
- Verdaguer B *et al.*, *Plant Mol Biol* 37:1055-1067, 1998.
- Waterhouse, *et al.*, *Proc. Natl. Acad. Sci. USA* 95:13959-13964, 1998.
- 15 Weigel D, *et al.*, *Plant Physiology*, 122:1003-1013, 2000.
- Wilson K *et al.*, *Plant Cell* 8: 659-671, 1996.
- Yadav NS *et al.*, (1993) *Plant Physiol* 103, 467-476.

## IT IS CLAIMED:

1. A transgenic plant comprising a plant transformation vector comprising a nucleotide sequence that encodes or is complementary to a sequence that encodes a HIO1005 polypeptide comprising the amino acid sequence of SEQ ID NO:3, or 4, or an ortholog thereof, whereby the transgenic plant has a high oil phenotype relative to control plants.  
5
2. The transgenic plant of claim 1, which is selected from the group consisting of rapeseed, soy, corn, sunflower, cotton, cocoa, safflower, oil palm, coconut palm, flax, castor and peanut.  
10
3. A plant part obtained from the plant according to claim 1.
4. The plant part of claim 3, which is a seed.  
15
5. Meal, feed, or food produced from the seed of claim 4.
6. A method of producing oil comprising growing the transgenic plant of claim 1 and recovering oil from said plant.  
20
7. The method of claim 6, wherein the oil is recovered from a seed of the plant.
8. A method of producing a high oil phenotype in a plant, said method comprising:
  - a) introducing into progenitor cells of the plant a plant transformation vector comprising a nucleotide sequence that encodes or is complementary to a sequence that encodes a HIO1005 polypeptide comprising the amino acid sequence of SEQ ID NO:3 or 4, or an ortholog thereof, and  
25
  - b) growing the transformed progenitor cells to produce a transgenic plant, wherein said polynucleotide sequence is expressed, and said transgenic plant exhibits an altered oil content phenotype relative to control plants.  
30
9. A plant obtained by a method of claim 8.
10. The plant of claim 9, which is selected from the group consisting of rapeseed, soy, corn, sunflower, cotton, cocoa, safflower, oil palm, coconut palm, flax, castor and peanut.  
35

11. A method of generating a plant having a high oil phenotype comprising identifying a plant that has an allele in its HIO1005 gene that results in increased oil content compared to plants lacking the allele and generating progeny of said identified plant, wherein the  
5 generated progeny inherit the allele and have the high oil phenotype.
12. The method of claim 11 that employs candidate gene/QTL methodology.
13. The method of claim 11 that employs TILLING methodology.  
10
14. A feed, meal, grain, food, or seed comprising a polypeptide encoded by the nucleic acid sequence as set forth in SEQ ID NO:1 or 2.
15. A feed, meal, grain, food, or seed comprising a polypeptide comprising an amino acid  
15 sequence as set forth in SEQ ID NO:3 or 4, or an ortholog thereof.



## SEQUENCE LISTING

<110> Agrinomics, LLC  
 Davies, John  
 Peters, Sandra  
 Ng, Hein T.

<120> GENERATION OF PLANTS WITH ALTERED OIL CONTENT

<130> 6616-71240-02

<150> 60/575,561

<151> 2004-05-28

<160> 12

<170> PatentIn version 3.3

<210> 1

<211> 1572

<212> DNA

<213> Arabidopsis thaliana

<400> 1

```

atgattact agccagctca attttctttc tttctttctt tctttctctg ggagtttggg      60
attttcttaa gggctttttt caattctcgc tccatthttgt acgcatagaa ctgggtttgt    120
gtttacacac ataaagctct tatctttttt cttctttcaa tcgagaagat tatagagctg    180
agtgactgat tgggtggtttc ttggagtttt aatgggaggt tgtttcagca atcggattaa    240
aacagatatt gcttccagta catggctaag ttcgaaatc ttgagtagag atgggagcaa    300
gggctcgtcg accgcttctt tctcttatat gcctcgaaca gaaggcgaga tcttgcaaaa    360
tgctaacttc aagaacttta gtctcagtga actgaaatct gcaactagga atttccggcc    420
tgatagtgtg gttgggtgaag gtggatttgg ttgcgthttc aaaggctgga tcgatgagtc    480
ctctctcgtc ctttctaaac cggggaccgg gattgtcatt gctgtgaaaa gacttaacca    540
agaagggttt caaggtcatc gagagtggct ggctgagatc aattatthtag gccagctgga    600
tcatcctaac cttgtgaaac tgattggata ctgcttgga gaggagcaca ggcttcttgt    660
ttacgagttt atgactcgtg gtagtcttga gaatcactta ttcagaagag gaacattcta    720
tcagccactt tcatggaaca cgcgggttcg tatggctctt ggtgcagcta gaggacttgc    780
ttttcttcac aatgctcaac cgcaagttat ataccgagac ttcaaagcat ctaacatctt    840
gctagattcg aactacaacg caaagctttc ggatttccgg ttggctagag atggtccaat    900
gggtgacaac agccatgtht ctaccagagt catgggaact cagggatacg ctgctccaga    960
atatctagct acaggtcatt tatcggtgaa gagcagtgta tacagthttg ggttgtgtgt  1020
actggagttg ttatcaggaa gacgagcaat tgacaagaat caaccagtag gagaacacaa  1080
tctcgtggat tgggcaagac cctacttaac aaacaagaga agacttctgc gagtgatgga  1140

```

tcctcgtctc caaggtcaat actcactaac ccgagctttg aaaattgcag ttcttgcact 1200  
 cgattgcata tctatagatg ccaagagtag accgacccatg aacgaaatcg tcaagacaat 1260  
 ggaagaactt catatccaga aggaagcatc aaaagagcag cagaatcctc aaatcagcat 1320  
 tgacaacatc atcaacaaat ctccacaagc tgtgaattat cctaggcctt caattatgta 1380  
 acaatcctag gcgagctatt taccgggttt tagagatgta tagactcttt accttctgtc 1440  
 tgtttagata ttatgttggt tggtagtaac aaaagagctg gcaatgtaag ggagagaagg 1500  
 aaacttacta gttgtaaact taggttctct tacaacggtc acatgttatc tcacatacaa 1560  
 aatgttatca gg 1572

<210> 2

<211> 1570

<212> DNA

<213> *Arabidopsis thaliana*

<400> 2

acgatgaaca gaggaagatg aggaagaaga caaagtccaa tagggcgtac caaaaactga 60  
 gagagaaaga gagaggaaag agctcaaaaa aaaacatctt ttcttttcat tttcaattag 120  
 tttcattatt atcacaaaag acttttcttc tccgagagaa caaaagatta tagagctgag 180  
 tgactgattg gtggtttctt ggagtttaa tgggaggtt tttcagcaat cggattaana 240  
 cagatattgc ttccagtaca tggctaagtt cgaaattctt gagtagagat gggagcaagg 300  
 gctcgtcgac cgcttccttc tcttatatgc ctccgaacaga aggcgagatc ttgcaaaatg 360  
 ctaatctcaa gaactttagt ctcaagtgaac tgaatctgc aactaggaat ttccggcctg 420  
 atagtgtggt tggggaaggt ggatttggtt gcggtttcaa aggctggatc gatgagtcct 480  
 ctctcgctcc ttctaaaccg gggaccggga ttgtcattgc tgtgaaaaga cttaccaag 540  
 aagggtttca aggtcatcga gagtggctgg ctgagatcaa ttatttaggc cagctggatc 600  
 atcctaacct tgtgaaactg attggatact gcttgaaga ggagcacagg cttcttgttt 660  
 acgagtttat gactcgtggt agtcttgaga atcaattatt cagaagagga acattctatc 720  
 agccactttc atggaacacg cgggttogta tggctcttgg tgcagctaga ggacttgctt 780  
 ttcttcacaa tgctcaaccg caagttatat accgagactt caaagcatct aacatcttgc 840  
 tagattcgaa ctacaacgca aagctttcgg atttcggttt ggctagagat ggtccaatgg 900  
 gtgacaacag ccatgtttct accagagtca tgggaactca gggatacgtc gctccagaat 960  
 atctagctac aggtcattta tccgtgaaga gcgatgtata cagttttggg gttgtgttac 1020  
 tggagttggt atcaggaaga cgagcaattg acaagaatca accagtagga gaacacaatc 1080  
 tcgtggattg ggcaagaccc tacttaacaa acaagagaag acttctgcga gtgatggatc 1140

ctcgtctcca aggtcaatac tcaactaacc gagctttgaa aattgcagtt cttgcactcg 1200  
 attgcatatc tatagatgcc aagagtagac cgaccatgaa cgaaatcgtc aagacaatgg 1260  
 aagaacttca tatccagaag gaagcatcaa aagagcagca gaatcctcaa atcagcattg 1320  
 acaacatcat caacaaatct ccacaagctg tgaattatcc taggccttca attatgtaac 1380  
 aatcctaggg gagctattta ccgggtttta gagatgtata gactctttac cttctgtctg 1440  
 tttagatatt atggtgtttg gtagtaacaa aagagctggc aatgtaaggg agagaaggaa 1500  
 acttactagt tgtaaactta ggttctctta caacgttcac atggttatctc acatacaaaa 1560  
 tgttatcagg 1570

<210> 3  
 <211> 389  
 <212> PRT  
 <213> Arabidopsis thaliana

<400> 3

Met Gly Gly Cys Phe Ser Asn Arg Ile Lys Thr Asp Ile Ala Ser Ser  
 1 5 10 15  
 Thr Trp Leu Ser Ser Lys Phe Leu Ser Arg Asp Gly Ser Lys Gly Ser  
 20 25 30  
 Ser Thr Ala Ser Phe Ser Tyr Met Pro Arg Thr Glu Gly Glu Ile Leu  
 35 40 45  
 Gln Asn Ala Asn Leu Lys Asn Phe Ser Leu Ser Glu Leu Lys Ser Ala  
 50 55 60  
 Thr Arg Asn Phe Arg Pro Asp Ser Val Val Gly Glu Gly Gly Phe Gly  
 65 70 75 80  
 Cys Val Phe Lys Gly Trp Ile Asp Glu Ser Ser Leu Ala Pro Ser Lys  
 85 90 95  
 Pro Gly Thr Gly Ile Val Ile Ala Val Lys Arg Leu Asn Gln Glu Gly  
 100 105 110  
 Phe Gln Gly His Arg Glu Trp Leu Ala Glu Ile Asn Tyr Leu Gly Gln  
 115 120 125  
 Leu Asp His Pro Asn Leu Val Lys Leu Ile Gly Tyr Cys Leu Glu Glu  
 130 135 140

Glu His Arg Leu Leu Val Tyr Glu Phe Met Thr Arg Gly Ser Leu Glu  
 145 150 155 160

Asn His Leu Phe Arg Arg Gly Thr Phe Tyr Gln Pro Leu Ser Trp Asn  
 165 170 175

Thr Arg Val Arg Met Ala Leu Gly Ala Ala Arg Gly Leu Ala Phe Leu  
 180 185 190

His Asn Ala Gln Pro Gln Val Ile Tyr Arg Asp Phe Lys Ala Ser Asn  
 195 200 205

Ile Leu Leu Asp Ser Asn Tyr Asn Ala Lys Leu Ser Asp Phe Gly Leu  
 210 215 220

Ala Arg Asp Gly Pro Met Gly Asp Asn Ser His Val Ser Thr Arg Val  
 225 230 235 240

Met Gly Thr Gln Gly Tyr Ala Ala Pro Glu Tyr Leu Ala Thr Gly His  
 245 250 255

Leu Ser Val Lys Ser Asp Val Tyr Ser Phe Gly Val Val Leu Leu Glu  
 260 265 270

Leu Leu Ser Gly Arg Arg Ala Ile Asp Lys Asn Gln Pro Val Gly Glu  
 275 280 285

His Asn Leu Val Asp Trp Ala Arg Pro Tyr Leu Thr Asn Lys Arg Arg  
 290 295 300

Leu Leu Arg Val Met Asp Pro Arg Leu Gln Gly Gln Tyr Ser Leu Thr  
 305 310 315 320

Arg Ala Leu Lys Ile Ala Val Leu Ala Leu Asp Cys Ile Ser Ile Asp.  
 325 330 335

Ala Lys Ser Arg Pro Thr Met Asn Glu Ile Val Lys Thr Met Glu Glu  
 340 345 350

Leu His Ile Gln Lys Glu Ala Ser Lys Glu Gln Gln Asn Pro Gln Ile  
 355 360 365

Ser Ile Asp Asn Ile Ile Asn Lys Ser Pro Gln Ala Val Asn Tyr Pro  
 370 375 380

Arg Pro Ser Ile Met

385

<210> 4  
 <211> 389  
 <212> PRT  
 <213> Arabidopsis thaliana

&lt;400&gt; 4

Met Gly Gly Cys Phe Ser Asn Arg Ile Lys Thr Asp Ile Ala Ser Ser  
 1 5 10 15

Thr Trp Leu Ser Ser Lys Phe Leu Ser Arg Asp Gly Ser Lys Gly Ser  
 20 25 30

Ser Thr Ala Ser Phe Ser Tyr Met Pro Arg Thr Glu Gly Glu Ile Leu  
 35 40 45

Gln Asn Ala Asn Leu Lys Asn Phe Ser Leu Ser Glu Leu Lys Ser Ala  
 50 55 60

Thr Arg Asn Phe Arg Pro Asp Ser Val Val Gly Glu Gly Gly Phe Gly  
 65 70 75 80

Cys Val Phe Lys Gly Trp Ile Asp Glu Ser Ser Leu Ala Pro Ser Lys  
 85 90 95

Pro Gly Thr Gly Ile Val Ile Ala Val Lys Arg Leu Asn Gln Glu Gly  
 100 105 110

Phe Gln Gly His Arg Glu Trp Leu Ala Glu Ile Asn Tyr Leu Gly Gln  
 115 120 125

Leu Asp His Pro Asn Leu Val Lys Leu Ile Gly Tyr Cys Leu Glu Glu  
 130 135 140

Glu His Arg Leu Leu Val Tyr Glu Phe Met Thr Arg Gly Ser Leu Glu  
 145 150 155 160

Asn His Leu Phe Arg Arg Gly Thr Phe Tyr Gln Pro Leu Ser Trp Asn  
 165 170 175

Thr Arg Val Arg Met Ala Leu Gly Ala Ala Arg Gly Leu Ala Phe Leu  
 180 185 190

His Asn Ala Gln Pro Gln Val Ile Tyr Arg Asp Phe Lys Ala Ser Asn  
 195 200 205

Ile Leu Leu Asp Ser Asn Tyr Asn Ala Lys Leu Ser Asp Phe Gly Leu  
 210 215 220

Ala Arg Asp Gly Pro Met Gly Asp Asn Ser His Val Ser Thr Arg Val  
 225 230 235 240

Met Gly Thr Gln Gly Tyr Ala Ala Pro Glu Tyr Leu Ala Thr Gly His  
 245 250 255

Leu Ser Val Lys Ser Asp Val Tyr Ser Phe Gly Val Val Leu Leu Glu  
 260 265 270

Leu Leu Ser Gly Arg Arg Ala Ile Asp Lys Asn Gln Pro Val Gly Glu  
 275 280 285

His Asn Leu Val Asp Trp Ala Arg Pro Tyr Leu Thr Asn Lys Arg Arg  
 290 295 300

Leu Leu Arg Val Met Asp Pro Arg Leu Gln Gly Gln Tyr Ser Leu Thr  
 305 310 315 320

Arg Ala Leu Lys Ile Ala Val Leu Ala Leu Asp Cys Ile Ser Ile Asp  
 325 330 335

Ala Lys Ser Arg Pro Thr Met Asn Glu Ile Val Lys Thr Met Glu Glu  
 340 345 350

Leu His Ile Gln Lys Glu Ala Ser Lys Glu Gln Gln Asn Pro Gln Ile  
 355 360 365

Ser Ile Asp Asn Ile Ile Asn Lys Ser Pro Gln Ala Val Asn Tyr Pro  
 370 375 380

Arg Pro Ser Ile Met  
 385

<210> 5  
 <211> 2205  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> synthetic construct

<400> 5  
 gtttatactt tatagaaact atggaaccaa gcaaaggcat ttgctttgaa ggccttaaac 60  
 tatattaataaa agccataatc atttcctaaa tatgccttgc ttaattagcg agaacttagc 120

aacacagaga	caaaaatcca	acagctaagc	aacaacagat	aaagttgaaa	ctaagttaag	180
ccattaacaa	tcttttgatt	tgattaagta	gttaatcccc	agttaatfff	tcagtctctt	240
ttggttgctt	gtaatctgtg	ctgctactag	taaagacaga	gaaccaacaa	agcaagtcaa	300
ccaaccatct	aatctatggc	tttttcagtt	ttcaccaaca	aaaagtccaa	aaaaaacttc	360
ataagtacca	tagagaatca	agaaaagcaa	taaaaccctt	ttttcctctt	gtccttttac	420
ttccaacttt	cttgtcataa	tttttttcat	tcataggcaa	gtgggggtaa	acaattacag	480
ccctgggcat	cagaaagttt	caatcttttt	cttttatctc	caaagttact	gccatttcat	540
gatgtagagg	aagattcttg	ggttcacact	agatctgggg	ttcaatgaac	ttttgggggt	600
tgtattgaag	taacaactgg	tgtaaatctg	tgcatataat	gatgaaaact	tgagttttgc	660
agtttgctg	gtggctagtg	gagctaaaaa	agattgaatt	ttggggtttt	tttttgctct	720
aatggggatt	tgtctgagta	accaaataca	ggctgagaca	acattttata	ctgtttcagg	780
gttgattca	agaaatgtca	gtggaaatgg	taccgatatt	agtaattcga	atagcaaaca	840
ctcatcagct	tccgtacccc	caactcctcg	gagtgagggt	gagatcttgc	agtcatacaa	900
cttgaggagc	ttcactttca	gtgaactccg	atctgcaact	agaaatttcc	gtcctgacag	960
tgtagtagga	gaaggagggt	ttggctcggg	tttcaaaggg	tgggtcgacg	agcatactct	1020
tgagcatca	aagcctggca	cagggattgt	gatagctgtg	aaaaagttaa	accaagaagg	1080
gtggcagggg	cacagagaat	ggctggctga	gataaactat	cttgggcaac	tgaccatccc	1140
aaatcttgta	aatttgatcg	gttattgctt	agaggaagat	cacaggctct	ttggcctatga	1200
gtttatgcct	aagggtagca	tggagaatca	tctatrttagg	agaggttcat	tctaccaacc	1260
actttcttgg	agccttcgta	tgaaagttgc	acttgggtgct	gcaaggggccc	ttgcgtttct	1320
tcataatgct	gaaacaaaag	ttatrttacag	ggacttcaag	acttctaata	ttctgctgga	1380
ctcggactac	aatgccaagc	tttctgattt	tgggttggcc	agagatggtc	ccatagggtga	1440
tcagagccac	gtgtctactc	gggttatggg	aacttatggc	tatgctgctc	cagagtatct	1500
atccacaggc	catcttactg	ccaagagtga	tgtatacagc	ttcggagttg	ttctcttgga	1560
aattctatca	ggtaagaaaag	caatagacaa	gaatcgacca	acaggggagc	acaatcttgt	1620
cgagtgtca	agaccttact	tgaccagtaa	acgtagagtt	ttccgtgttc	tagattctcg	1680
acttgaagga	caatactcac	tactcgtgc	ccttaaggta	gccaacgttg	cacttcaatg	1740
cctagccatg	gacccaagt	caagaccaac	tatggatgag	gtagtaacag	ctctagagca	1800
gctacaagag	tccaaggatc	gcgtgaaaaa	tgataaaaac	aaggatcaac	agttgaaatcg	1860
acttagcagt	caatcaagtg	gcgagctcaa	caagtcattc	agaagcaatt	cagaagagac	1920

tccgcgtgta gccaaattatc ccagaccttc agcttccctt cgctctatct gaaaagacgg 1980  
 gaattcacat tcttggcgtg ccagttgta acatccaatt ttaaggaact tctggcctgc 2040  
 aagtactact acaacacgaa taacatatcc agtgtagtcc cacaaaatgg gatatgtaag 2100  
 tgtaccaata taatcaacat gtatagtttt ctttttttga tgtttacttt gtctaataac 2160  
 tgttgaaatg aggattgatg agatgatgta taatggattg caagt 2205

<210> 6  
 <211> 1601  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> synthetic construct

<400> 6  
 ttttgataaa agttttcttt cttataaaga ttgaaacttt ggttcataaa ttttaatggg 60  
 ttcttgttta agtgttcgaa ttaaagcaga aagtcctctt catcatggag caagtgatgg 120  
 aagagagttg agtagtaggc attcatattc atcagctccg ttgactcctc gtagtcagag 180  
 tgaaataactt gaatcgtcga atctgaaaag ttttagtttc aatgaactcc gagtagccac 240  
 gaggaacttt cgtccagata gtgtgttggg agaaggtggg tttggttgtg tatttaaggg 300  
 ttggatagac gagaatacgt ttaaagctgc aaggccaggg actggattgg ttatagctgt 360  
 aaagagattg aaccaagaaa gctttcaagg tcacaaggaa tggctggcgg aaattactta 420  
 cctgagccag ctatctcatc ctaatctcgt gaagttgatt ggatattgct tagaagatga 480  
 acacaggctt ctggtgtacg aattcatgcc aagaggaagt ttggaaaatc atctattcag 540  
 aaggagtact tatttccaac cactatcatg gaatctccga atgaaggttg ctctcgaggc 600  
 agctaagga ctgacatata tccacagtcc agaagctaaa gttatatatac gggatttcaa 660  
 atcgtcgaac attttgcttg atgctaatta caatgcaaag ctttctgatt ttggattggc 720  
 gaaggatggg ccaatcgacg gtaaaagcca tgtctctact agagtaatgg gtacctttgg 780  
 ttatgctgct cccgagtata tggccacagg tcatataacc actagaagtg acgtatacag 840  
 tttcgggggt gttcttctag aatgctgac aggccgtcga gtgatggaca aaaaccgacc 900  
 ccacggggaa cataatctga ttgaatgggc taaacctttc cttactagta aacgtaaagt 960  
 cctccatatt atggatcacc gtatagaagg tcagtactca gtggaaggag cactgagagc 1020  
 agcacttctt gcagtcaaat gcctggcatt agaaccctaaa ttcaggccta aatgcccga 1080  
 ggtagtaaag gcattggagc aacttcagaa cttgaatgaa tcaggaagtc tgaaacgcga 1140  
 aaccacacaa aggaagcatc acagaacctc aaccgatgaa gcttcaggaa gaaaaactac 1200  
 gtcttatccg aggccagctg cttctcccct tgttacatga atggagaatg aatggtaaca 1260



tctttaacca tatacacact acctgaaccg aaaggaaacg atgagacgac ttacttggtg 1320  
aagatatatc tatatctgct gatagagcac gaagagtaat ttgatcaaat tcgcgacat 1380  
tttatgagaa ggtatgctta tgatactaga ctgaatcttg gaatctagga actatggtgt 1440  
aaatttgat caattatggt agcaatctac agttcttatt ttgatctagc tggaaattta 1500  
acattgatgt tttctcaca catctatgct aaattgactt ttgttgatca tcttttgaga 1560  
gccaaagatg tgaaaaggac ttgtaactt atttgactg c 1601

<210> 7  
<211> 965  
<212> DNA  
<213> Artificial sequence

<220>  
<223> synthetic construct

<400> 7  
caactgcggg aacgcgagag gagaatgccg tcgtgcggc gcacgccca gttcagcagc 60  
tccatttggt acaacatcct gtcaagaatg ctggtgcaga gaggaagcat acccgcatct 120  
catcagatat gagtgatcct tcaacaccta ggaaaattga agatgccaa aacatctcca 180  
tatacaacga tgtgattgat ttcacattgt ttgaactcga gaccatcaca aagagcttcc 240  
gtgctgacta cgttcttggg gaaggagggt ttgggactgt ttacaagggc tacatagatg 300  
agaatgtcag ggttggctct aagtcactac ctggtgcagt caagggtgct aacaaagatg 360  
gacatcaagg acacagagaa tggcttactg aggttaggtt ccttgggcag ctaagacatc 420  
caaatttagt caagttgatt gggattgct gcgaagatga ccacaggctg cttgtgtacg 480  
agttcatggt tcgaggaagt ctagaaaacc acttattccg aaggacagct actccactat 540  
cctgggctac taggatgtcg attgcattag gggctgcca agggttagct tgcctccaca 600  
atgctgaaag gccaaattatc tacagggatt tcaagacatc aaatattctg ctggactcag 660  
attatactgc taaactctct gactttggc tggcaaaagc tggcccagaa ggcgatcaaa 720  
cccatgtatc aacacgggtg atgggaacat atggttatgc tgcccctgaa tacgtgatga 780  
ctggctcact gactgctaga agtgatgtct acagcttggg tgttgcctt cttgaactct 840  
tgaactgggc gtaagtccat cgacaagtca cggcccagta gggagcacag cttggttgac 900  
tgggccctcc tgaaactgaa cgacaagagg aggttctctc aaatcattga cccaaaactg 960  
gaggg 965

<210> 8  
<211> 1242  
<212> DNA

<213> Artificial sequence

<220>

<223> synthetic construct

<220>

<221> misc\_feature

<222> (1207)..(1207)

<223> n is a, c, g, or t

<220>

<221> misc\_feature

<222> (1216)..(1216)

<223> n is a, c, g, or t

<400> 8

```

gtccagacac agtcacgcac caagcaagca ccacggggag ggactctgtc ttctcttttc      60
tattctcoct cctctcctct ccaccaacaa cgtcctttgt tttctattat taccactttc      120
aagaactagt ataatataat ttcttgcttc tccatcggta ttctatcaaa atattatagt      180
gatattttct tttttgtatt tgtacttgac cgaacccaaa aataaacagc ctccaagtct      240
tcgtctttat ctattataat gtgacatggg ggtttcttct tgatttctct caccaagaca      300
ccagaagaag aatttaggat tccaatagcg gttgctttgc tttggtttac caaatgggca      360
actgcttaga ttcttctgct gctaaagtag attccactca gagttcatat actcctgcat      420
caggagcctc aagaatttcc agcagaacca gccggttctc agttccttct agtctgacca      480
tcccatcata cagtgggaag agcagttctg aatgttttcc tacaccaagg agtgaagggtg      540
aaatattgtc atctccaaat ttaaaggcct tctcattcaa cgagctaaaa agtgccacca      600
gaaacttccg tcccagacagt cttctcggtg aagggtggctt tggttgtgtt ttcaaaggat      660
ggattgatga aaacacgttg actgcttcaa agcctggatc aggaatgggt gtggcagtca      720
agaagcttaa acctgaaggt ttccaaggcc acaaggagtg gttgacagaa gttaattatc      780
ttggccaact tcatcatcca aatctggtta aattgattgg gtactgctg gaaggtgaga      840
accgacttct ggtctatgag ttcattgccta aaggagactt ggagaatcat ctgttcagaa      900
gaggaccaca gccactttca tgggcagtaa ggggtcaaagt ggctataggt gctgccagag      960
ggctctcttt tcttcatgat gcgaaatcac aagtcataata ccgtgacttc aaggcatcta     1020
atattctact agatgcggaa ttaaatgcaa aactttctga ttttggctctg gccaaggcag     1080
gccctactgg tgataggacc cacgtgtcca ctcaagttat ggggtactcat gggtatgcag     1140
cacctgaata cgttgctaca ggtcggttga cagccaaaag tgatgtatac agctttgggg      1200
ttgtgtngct tgaatnattg tctgggcgac gggctgtaga ta                          1242

```

<210> 9

<211> 1631  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> synthetic construct

<400> 9  
 gatccttctc ctctcatcaa accccaaatg atagaaagct gaaaaaagg tttctgaaat 60  
 atcagcta at gttactgagt gcgtaatcat tgagatctgc cttcatcttg gggtgatttt 120  
 gttgatgggt gcttgctgga gtaataggat taaggctgtg agtccttcca atacagggat 180  
 cacttctaga agtgtcagca ggagtggcca tgacatcagc tcaaatagca ggagctcatc 240  
 agcctccata cccgtcactt ctcgagtgga gggtgagatc ttgcaatctt ccaacttgaa 300  
 aagcttcagc tatcatgagc taagagcagc cacaagaaat ttccgccccg atagtgtctt 360  
 gggagagggc gggtttggtt cagtttttaa gggctggatt gatgagcatt cacttgctgc 420  
 taccaaacca ggaataggca agattgttgc tgttaagaag cttaaccaag atgggctcca 480  
 gggtcacaga gagtggttgg ctgaaataaa ctatcttggc caactacagc atcctaacct 540  
 tgtcaagtta ataggatact gctttgagga tgaacatcgg cttctggttt atgaatttat 600  
 gcctaagggt agcatggaaa atcatctatt cagaagaggt tcttactttc agccattctc 660  
 ttggagtttg cgaatgaaaa tagcactagg ggctgcaaag ggtcttgctt ttcttcatag 720  
 tacagaacat aaagtcatat accgtgactt taaaacgtca aatatectac tcgatacaaa 780  
 ctataatgcc aaactttctg attttggtt ggccagagat ggtcctactg gtgataaaaag 840  
 ccatgtctct actagggtca tgggaaccgg tggatatgcg gcaccagagt atttagcgac 900  
 aggtcatctc actaccaaga gtgatgtgta tagttttgga gtagttcttc tggaaatgat 960  
 atcaggaaga cgagctattg acaagaacca gccaaactgga gagcataacc ttggtgaatg 1020  
 ggccaagcct tatctatcta ataaacgaag agtcttccgt gtaatggatc cccgtctcga 1080  
 aggtcaatat tcacagaatc gagctcaagc agcagctgcg cttgctatgc aatgtttttc 1140  
 cgtagagcct aagtgcaggc cgaatatgga tgagggtgta aaagcattgg aggagcttca 1200  
 ggaatcaaag aacatgcaga gaaaaggcgc tgatcataaa cagcatcatg tacgcaattc 1260  
 cggccccggc cgtaccaatg gtggcgatgg tggttcagat gccctagaa aggcttctgc 1320  
 ttatcctaga ccttctgctt ctctcctccg cggttgagag aattttgatga tgcggcaagt 1380  
 agtaaagcat ataacaaaat tcggaggttt aggtgtggct ggcataacc atgtttatgt 1440  
 aagcaatgta tagtttgctt tatttccatg ttacattttg tgctatata aatttttttt 1500  
 tctcgttggt gttttgattg ttagacattc aagggtttaa tgtgggtcct taagggttcc 1560  
 attaattttg aacctgtaa acaactgcaa cactacccat gtggttgatt ctctacattt 1620

ttatccttgc t 1631

<210> 10  
 <211> 1620  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> synthetic construct

<400> 10  
 tttttttttt ttttttgccg cggcgggtct ttaagaagc atttcttaca ctctcaagaa 60  
 cattcaaact ggaatcctgc ggtgaagtgg aaaaaagaa cttacccttt gaaccgctgg 120  
 tacatcgaag actaataaag tatgtaggta gagaccaag ttacaacgac gaagcaacta 180  
 gctgtacaca aagcaaggca atattatatt ctcttcacaa gccacttagc cttcaaacc 240  
 ctcatccggt ctctgttttt tactccccta aagtattaca ggaggatgct ccagactcca 300  
 cagaggtaga tggacatctt tttcagttag cccttggcct ttgaagttgt tctaatgcag 360  
 tcacaacctg atccatgcca ggctgcatc gtgcgtccat cgataagcat tgcagcgcca 420  
 gggttgccac cttctgagcc gcgggaaggg aatactggga acccagccgt gagtccagga 480  
 catgaatgac cctcctcttg ttggtgatgt aaggcttagc cactccacc agattgtgct 540  
 gaccgggagg gcggttcttg tccagagcgc gctgcccgga cagcagctcc agaagaacca 600  
 caccgtagct gtacacatcg ctcttcgctg tcaaattggc tgcgcaaga tattcagggg 660  
 cagcgtatcc ttgtgtcccc atgaccctag tagaaacatg gctcttatca ccgcttggac 720  
 cgtcctttgc caaacgaaa tcagacagt ttgcgttgta ctccgagtcg agaaggatat 780  
 tggaggtctt gaaatcgcgg tagataactt tggcctggtc gccgtgcagg aaagcaagac 840  
 ctctggcagc ctcgagagca acctcatgc gtaagttcca tgggagcggc tggagttgg 900  
 agcccctcct gaagagatga tgctccaggc tccctctcgg catgtactcg tacacaagaa 960  
 gccgctgctc gtcctccaag cagtagccga tgagtttgac gagattaggg tgggataact 1020  
 gcccaggta attgacctcc gccagccatt cccggtgccc ttggaagctg tgcagcttga 1080  
 gcttcttgac ggcgacgatg atcccggcgc ccggcctgac gggcgcgagc gtgcgctcat 1140  
 ccatccaccc cttgaacacg gaccgaagc cccctcccc gagcaggctg tccggcctga 1200  
 agttcctcgt ggagcctttg agctcggcga agctgaactt gcggacattc gacgactcca 1260  
 ggatctccgc ctccgtacga ggggttggcg gcaccgacga cgacgagccg gacgtcttgg 1320  
 tgccgaaggt gctggtgctg gccgcgttct tgetgacgct cttagaacg gcggtccctg 1380  
 aggaattggg ggcactgaag gaagggttga tctcggcgtt ccctccacg ccggcgcagt 1440

tccccatcag gaacgagctc tcgcggcagc gcgcagggcg ggggggagag ggaggtagaa 1500  
 gcggttagct gaactacgtc tgggtttatt tgccctcgtg tgtcatcctc tcgctgcgtg 1560  
 ggaaggaggt gatgagtaca atggctgctg cagtgcctgc ccggctggct ctggctctgg 1620

<210> 11  
 <211> 632  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> synthetic construct

<400> 11  
 gtgatttcaa agcttctaatt attctactag atgcggaatt caatgcaaaa ctttctgatt 60  
 tcggattggc caaggcaggg cctactggcg atagaactca tgtatcgact caggttatgg 120  
 gaactcaagg ctatgcagca cctgaatatg tcgccacagg tcgattgacg gctaagagcg 180  
 atgtttacag ctttggagtt gttttgctcg aactactgtc tggacagcgt gctgttgata 240  
 aaacaaaagt tggcgtggag cggagtctag taaactgggc aaaaccatat ttgggtgaca 300  
 aaagaaaatt attccgaatc atggacacga agttgggtgg ccagtacccc cagaaagggtg 360  
 cttttacggc agctaccctt gctttacagt gcctaaatag tgaagctaaa ctcaggcctc 420  
 gaatgtcaga agttttggca gactggagc agctagaagc ccccaaaact gcatcaaaac 480  
 acagtcaagc agaacagcat gctgttccac ttccctgttcg gagatcccca atgcgacacc 540  
 atcgatcacc tatgcatcta acaccagtg catctccatt gccatctcac agacaatccc 600  
 cgcggttaca ttgagattcc atctgtaaatt tc 632

<210> 12  
 <211> 1921  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> synthetic construct

<400> 12  
 attcggcagc agggatccca ccaccgccc ctctcattcg cagctccgtt cccacctgag 60  
 ccagccggt cgggtggctgc cggcagcaac ctgctgctac ttccgttgct cctcctccct 120  
 ccctcccac tccctctcag cccagctcg agcgagagag catctccctc ccgagcgacc 180  
 gcgtccgct cctcccctg cccccctgc ggttcttggg gacggcgcgg cggcttggt 240  
 ggccggcctc gcctccagtt gaccgactg actctcccc attcggtcag cccggaggag 300  
 gaggaggagg cggaggggat ttgcttgact ttttgaagct tctcgtcgt ttctctccta 360  
 gcgttgccgc tgccggccaa cggcctcgcg gatttttctc cagtctcacc ccctccttca 420

agtggcggcg gagatcagct ggatccaaaa gcctcgccag aaaccggctg ggactctagg 480  
 gttggctgct cagctcagtt gtcttcggcg gctgctcctc ctgatccctc ccggtgccgc 540  
 tggccagctc tctctggctt tgctgacttg acccgggacg ggggaaagag gagcagagtt 600  
 cgctgtgga tcttcgggaa aggaagetca ttcacctgt cagctgggag gtagaggaag 660  
 agatggggaa ctgcatgaag tccacggccc ggggtggatca cagcatgaac actagtgccg 720  
 cctatccatc gaaagtgacc agcaaaaca gcatgtcatc tgctacttct gcgagcaaga 780  
 ccaactcaac tcgttcaacg tttactctgc catctataag agatcgcagc gagccgcctc 840  
 ggacagaagg cgaaatcttg tcatcgtcga acttgaaggc cttcttgttc aatgatctta 900  
 aaaatgcgac caagaacttc cgcccggaca gtcttcttgg ggaaggaggg tttgggcatg 960  
 ttttcaaagg ttggattgat gaacacactc ttgctccttc gaaaccggga agcggtatgg 1020  
 ttgttgctgt caagaagctt aaaccagaag gtttccaagg gcacaaggaa tggctgacgg 1080  
 aggttaacta ccttggccaa cttcaccatg ccaatcttgt taagctcatt ggttattgct 1140  
 cagatggtga caacagactt ctggtgatg agttcatgcc caagggaggt ttggagaatc 1200  
 atctgttcag aagaggtgct gatcctttat catggggaat aaggcttaag gttgctatcg 1260  
 gggctgctaa gggtttgca tttttacatc atgctgaaaa ccaagttata taccgtgatt 1320  
 tcaaggcatc aaacattctt cttgactcgg aattcaacgc gaagctttca gattttggat 1380  
 tggcgaaaagc tggccaact ggggataaaa ctcatgtttc cacacaagtg atgggcaccc 1440  
 atggatatgc agctccagag tatatcgcaa caggctgcct ctctgctaag gcagacgtct 1500  
 acagcttcgg ggtggtggtg cttgagttgc tgaccgggag acgagccctg gacaagtcaa 1560  
 agccaggcat agagcagaac ctagtggact gggccaaacc gcacctgcgc gacaagcgca 1620  
 ggctgtaccg tgtcatggac acgaagctgg gagggcagta tcctaagaaa ggcgcacacg 1680  
 ccgtcgcgaa cctcgccttg caatgatct gcaacgacgc caagatgcgg ccgcagatct 1740  
 cagaggtctt agaagagctg gagcagctcc aagactccaa aagcaatata gtgtcgccgc 1800  
 aggttgacat ccggaggacc tccaacaccg tcccgaagtc gccaatgagg ggccaacct 1860  
 cgccgcggcg ctctcttga gcatggcgc ccgctcaccg gcttttagga ccgcaagt 1920  
 g 1921