

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
23 February 2006 (23.02.2006)

PCT

(10) International Publication Number  
WO 2006/020717 A2

(51) International Patent Classification:

C12N 15/82 (2006.01) C07K 14/415 (2006.01)  
C12N 15/29 (2006.01) A01H 5/00 (2006.01)

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.

(21) International Application Number:

PCT/US2005/028431

(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, LV, 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).

(22) International Filing Date: 11 August 2005 (11.08.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/600,579 11 August 2004 (11.08.2004) US

Declarations under Rule 4.17:

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations 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, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations

(71) Applicant (for all designated States except US): BASF PLANT SCIENCE GMBH [DE/DE]; Carl-Bosch-Strasse 38, 67056 Ludwigshafen (DE).

(72) Inventors; and

(75) Inventors/Applicants (for US only): HAERTEL, Heiko [DE/US]; 24 Chestnut Bluffs Lane, Durham, NC 27713 (US). GIBSON, Jermaine [US/US]; 5538 Hampstead Crossing Drive, Raleigh, NC 27612 (US). SHANK, Karin, J. [US/US]; 1220 Nikole Court, Raleigh, NC 27612 (US).

(74) Agent: WU, Hui-Ju; Connelly Bove Lodge & Hutz LLP, The Nemours Building, 1007 North Orange Street, P.O. Box 2207, Wilmington, DE 19899 (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,

Published:

- without international search report and to be republished upon receipt of that report

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

(54) Title: NUCLEIC ACIDS CONFERRING LIPID AND SUGAR ALTERATIONS IN PLANTS II

(57) Abstract: Isolated nucleic acids and proteins associated with lipid and sugar metabolism regulation are provided. In particular, lipid metabolism proteins (LMP) and encoding nucleic acids originating from *Physcomitrella patens*, *Brassica napus*, *Glycine max*, *Zea mays*, and *Oryza sativa* are provided. The nucleic acids and proteins are used in methods of producing transgenic plants and modulating levels of seed storage compounds. Preferably, the seed storage compounds are lipids, fatty acids, starches or seed storage proteins.

WO 2006/020717 A2

## **Nucleic acids conferring lipid and sugar alterations in plants II**

### **BACKGROUND OF THE INVENTION**

#### **Field of the Invention**

[0001] Described herein are inventions in the field of genetic engineering of plants, including isolated nucleic acid molecules encoding casein kinase I-like polypeptides to improve agronomic, horticultural and quality traits. This invention relates generally to nucleic acid sequences encoding proteins that are related to the presence of seed storage compounds in plants. More specifically, the present invention relates to casein kinase I-like nucleic acid sequences encoding sugar and lipid metabolism regulator proteins and the use of these sequences in transgenic plants. The invention further relates to methods of applying these novel plant polynucleotides and polypeptides to the identification and stimulation of plant growth and/or to the increase of yield and/or composition of seed storage compounds.

#### **Background Art**

[0002] The study and genetic manipulation of plants has a long history that began even before the famed studies of Gregor Mendel. In perfecting this science, scientists have accomplished modification of particular traits in plants ranging from potato tubers having increased starch content to oilseed plants such as canola and sunflower having increased or altered fatty acid content. With the increased consumption and use of plant oils, the modification of seed oil content and seed oil levels has become increasingly widespread (e.g. Töpfer et al. 1995, *Science* 268:681-686). Manipulation of biosynthetic pathways in transgenic plants provides a number of opportunities for molecular biologists and plant biochemists to affect plant metabolism giving rise to the production of specific higher-value products. The seed oil production or composition has been altered in numerous traditional oilseed plants such as soybean (U.S. Patent No. 5,955,650), canola (U.S. Patent No. 5,955,650), sunflower (U.S. Patent No. 6,084,164) and rapeseed (Töpfer et al. 1995, *Science* 268:681-686), and non-traditional oil seed plants such as tobacco (Cahoon et al. 1992, *Proc. Natl. Acad. Sci. USA* 89:11184-11188).

[0003] Plant seed oils comprise both neutral and polar lipids (see Table 1). The neutral lipids contain primarily triacylglycerol, which is the main storage lipid that accumulates in oil bodies in seeds. The polar lipids are mainly found in the various membranes of the seed cells, e.g. the endoplasmic reticulum, microsomal membranes, and the cell membrane. The neutral and

polar lipids contain several common fatty acids (see Table 2) and a range of less common fatty acids. The fatty acid composition of membrane lipids is highly regulated and only a select number of fatty acids are found in membrane lipids. On the other hand, a large number of unusual fatty acids can be incorporated into the neutral storage lipids in seeds of many plant species (Van de Loo F.J. et al., 1993, *Unusual Fatty Acids in Lipid Metabolism in Plants*, pp. 91-126, editor TS Moore Jr. CRC Press; Millar et al. 2000, *Trends Plant Sci.* 5:95-101).

**[0004]** Lipids are synthesized from fatty acids and their synthesis may be divided into two parts: the prokaryotic pathway and the eukaryotic pathway (Browse et al. 1986, *Biochemical J.* 235:25-31; Ohlrogge & Browse 1995, *Plant Cell* 7:957-970). The prokaryotic pathway is located in plastids that are the primary site of fatty acid biosynthesis. Fatty acid synthesis begins with the conversion of acetyl-CoA to malonyl-CoA by acetyl-CoA carboxylase (ACCase). Malonyl-CoA is converted to malonyl-acyl carrier protein (ACP) by the malonyl-CoA:ACP transacylase. The enzyme beta-keto-acyl-ACP-synthase III (KAS III) catalyzes a condensation reaction, in which the acyl group from acetyl-CoA is transferred to malonyl-ACP to form 3-ketobutyryl-ACP. In a subsequent series of condensation, reduction and dehydration reactions the nascent fatty acid chain on the ACP cofactor is elongated by the step-by-step addition (condensation) of two carbon atoms donated by malonyl-ACP until a 16- or 18-carbon saturated fatty acid chain is formed. The plastidial delta-9 acyl-ACP desaturase introduces the first unsaturated double bond into the fatty acid. Thioesterases cleave the fatty acids from the ACP cofactor and free fatty acids are exported to the cytoplasm where they participate as fatty acyl-CoA esters in the eukaryotic pathway. In this pathway the fatty acids are esterified by glycerol-3-phosphate acyltransferase and lysophosphatidic acid acyl-transferase to the sn-1 and sn-2 positions of glycerol-3-phosphate, respectively, to yield phosphatidic acid (PA). The PA is the precursor for other polar and neutral lipids, the latter being formed in the Kennedy pathway (Voelker 1996, *Genetic Engineering ed.:Setlow* 18:111-113; Shanklin & Cahoon 1998, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49:611-641; Frentzen 1998, *Lipids* 100:161-166; Millar et al. 2000, *Trends Plant Sci.* 5:95-101).

**[0005]** Storage lipids in seeds are synthesized from carbohydrate-derived precursors. Plants have a complete glycolytic pathway in the cytosol (Plaxton 1996, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47:185-214) and it has been shown that a complete pathway also exists in the plastids of rapeseeds (Kang & Rawsthorne 1994, *Plant J.* 6:795-805). Sucrose is the primary source of carbon and energy, transported from the leaves into the developing seeds. During the storage phase of seeds, sucrose is converted in the cytosol to provide the metabolic

precursors glucose-6-phosphate and pyruvate. These are transported into the plastids and converted into acetyl-CoA that serves as the primary precursor for the synthesis of fatty acids. Acetyl-CoA in the plastids is the central precursor for lipid biosynthesis. Acetyl-CoA can be formed in the plastids by different reactions and the exact contribution of each reaction is still being debated (Ohlrogge & Browse 1995, *Plant Cell* 7:957-970). It is, however, accepted that a large part of the acetyl-CoA is derived from glucose-6-phosphate and pyruvate that are imported from the cytoplasm into the plastids. Sucrose is produced in the source organs (leaves, or anywhere that photosynthesis occurs) and is transported to the developing seeds that are also termed sink organs. In the developing seeds, sucrose is the precursor for all the storage compounds, i.e. starch, lipids, and partly the seed storage proteins. Therefore, it is clear that carbohydrate metabolism, in which sucrose plays a central role, is very important to the accumulation of seed storage compounds.

[0006] Storage compounds such as triacylglycerols (seed oil) serve as carbon and energy reserves, which are used during germination and growth of the young seedling. Seed (vegetable) oil is also an essential component of the human diet and a valuable commodity providing feed stocks for the chemical industry.

[0007] Although the lipid and fatty acid content and/or composition of seed oil can be modified by the traditional methods of plant breeding, the advent of recombinant DNA technology has allowed for easier manipulation of the seed oil content of a plant, and, in some cases, has allowed for the alteration of seed oils in ways that could not be accomplished by breeding alone (see, e.g., Töpfer et al., 1995, *Science* 268:681-686). For example, introduction of a  $\Delta^{12}$ -hydroxylase nucleic acid sequence into transgenic tobacco resulted in the introduction of a novel fatty acid, ricinoleic acid, into the tobacco seed oil (Van de Loo et al., 1995, *Proc. Natl. Acad. Sci USA* 92:6743-6747). Tobacco plants have also been engineered to produce low levels of petroselinic acid by the introduction and expression of an acyl-ACP desaturase from coriander (Cahoon et al., 1992, *Proc. Natl. Acad. Sci USA* 89:11184-11188).

[0008] The modification of seed oil content in plants has significant medical, nutritional and economic ramifications. With regard to the medical ramifications, the long chain fatty acids (C18 and longer) found in many seed oils have been linked to reductions in hypercholesterolemia and other clinical disorders related to coronary heart disease (Brenner, 1976, *Adv. Exp. Med. Biol.* 83:85-101). Therefore, consumption of a plant having increased levels of these types of fatty acids may reduce the risk of heart disease. Enhanced levels of seed

oil content also increase large-scale production of seed oils and thereby reduce the cost of these oils.

[0009] In order to increase or alter the levels of compounds such as seed oils in plants, nucleic acid sequences and proteins regulating lipid and fatty acid metabolism must be identified. As mentioned earlier, several desaturase nucleic acids such as the  $\Delta^6$ -desaturase nucleic acid,  $\Delta^{12}$ -desaturase nucleic acid and acyl-ACP desaturase nucleic acid have been cloned and demonstrated to encode enzymes required for fatty acid synthesis in various plant species. Oleosin nucleic acid sequences from such different species as canola, soybean, carrot, pine, and *Arabidopsis thaliana* have also been cloned and determined to encode proteins associated with the phospholipid monolayer membrane of oil bodies in those plants.

[0010] It has also been determined that two phytohormones, gibberellic acid (GA) and abscisic acid (ABA), are involved in overall regulatory processes in seed development (e.g. Ritchie & Gilroy, 1998, *Plant Physiol.* 116:765-776; Arenas-Huertero et al., 2000, *Genes Dev.* 14:2085-2096). Both the GA and ABA pathways are affected by okadaic acid, a protein phosphatase inhibitor (Kuo et al., 1996, *Plant Cell.* 8:259-269). The regulation of protein phosphorylation by kinases and phosphatases is accepted as a universal mechanism of cellular control (Cohen, 1992, *Trends Biochem. Sci.* 17:408-413). Likewise, the plant hormones ethylene (e.g. Zhou et al., 1998, *Proc. Natl. Acad. Sci. USA* 95:10294-10299; Beaudoin et al., 2000, *Plant Cell* 2000:1103-1115) and auxin (e.g. Colon-Carmona et al., 2000, *Plant Physiol.* 124:1728-1738) are involved in controlling plant development as well.

[0011] Although several compounds are known that generally affect plant and seed development, there is a clear need to specifically identify factors that are more specific for the developmental regulation of storage compound accumulation and to identify genes that have the capacity to confer altered or increased oil production to its host plant and to other plant species. This invention discloses nucleic acid sequences from *Physcomitrella patens*, *Brassica napus*, *Glycine max*, *Zea mays*, and *Oryza sativa*. These nucleic acid sequences can be used to alter or increase the levels of seed storage compounds such as proteins, sugars, and oils, in plants, including transgenic plants, such as Arabidopsis, canola, linseed, soybean, sunflower, maize, oat, rye, barley, wheat, rice, pepper, tagetes, cotton, oil palm, coconut palm, flax, castor, and peanut, which are oilseed plants containing high amounts of lipid compounds.

## SUMMARY OF THE INVENTION

[0012] The present invention provides novel isolated nucleic acid and amino acid sequences associated with the metabolism of seed storage compounds in plants, in particular with sequences that are casein kinase I-like.

[0013] The present invention also provides an isolated nucleic acid from *Physcomitrella patens*, *Brassica napus*, *Glycine max*, *Zea mays* and *Oryza sativa* encoding a Lipid Metabolism Protein (LMP), or a portion thereof, that are casein kinase I-like. These sequences may be used to modify or increase lipids and fatty acids, cofactors and enzymes in microorganisms and plants.

[0014] Mosses and algae are known systems that produce considerable amounts of fatty acids like linolenic acid and /or arachidonic acid and/or eicosapentaenoic acid and/or docosaheanoic acid (see, e.g., Table 2). Therefore, nucleic acid molecules originating from the moss *Physcomitrella patens* are especially suited to modify the lipid and fatty acid metabolism in a host, especially in microorganisms and plants like *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa*. Furthermore, nucleic acids from the moss *Physcomitrella patens* can be used to identify those DNA sequences and enzymes in other species like *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa*, which are useful to modify the biosynthesis of precursor molecules of fatty acids in the respective organisms.

[0015] The present invention further provides an isolated nucleic acid comprising a fragment of at least 15 nucleotides of a nucleic acid from a moss (*Physcomitrella patens*) or a plant (*Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa*) encoding a Lipid Metabolism Protein (LMP), or a portion thereof.

[0016] Also provided by the present invention are polypeptides encoded by the nucleic acids, and heterologous polypeptides comprising polypeptides encoded by the nucleic acids, and antibodies to those polypeptides.

[0017] Additionally, the present invention relates to and provides the use of LMP nucleic acids in the production of transgenic plants having a modified level or composition of a seed storage compound. In regard to an altered composition, the present invention can be used to, for example, increase the percentage of oleic acid relative to other plant oils. A method of producing a transgenic plant with a modified level or composition of a seed storage compound includes the steps of transforming a plant cell with an expression vector comprising a LMP

nucleic acid, and generating a plant with a modified level or composition of the seed storage compound from the plant cell. In a preferred embodiment, the plant is an oil producing species selected from the group consisting of canola, linseed, soybean, sunflower, maize, oat, rye, barley, wheat, rice, pepper, tagetes, cotton, oil palm, coconut palm, flax, castor, and peanut, for example.

[0018] According to the present invention, the compositions and methods described herein can be used to alter the composition of a LMP in a transgenic plant and to increase or decrease the level of a LMP in a transgenic plant comprising increasing or decreasing the expression of a LMP nucleic acid in the plant. Increased or decreased expression of the LMP nucleic acid can be achieved through in vivo mutagenesis of the LMP nucleic acid. The present invention can also be used to increase or decrease the level of a lipid in a seed oil, to increase or decrease the level of a fatty acid in a seed oil, or to increase or decrease the level of a starch in a seed or plant.

[0019] Also included herein is a seed produced by a transgenic plant transformed by a LMP DNA sequence, wherein the seed contains the LMP DNA sequence, and wherein the plant is true breeding for a modified level of a seed storage compound. The present invention additionally includes a seed oil produced by the aforementioned seed.

[0020] Further provided by the present invention are vectors comprising the nucleic acids, host cells containing the vectors, and descendent plant materials produced by transforming a plant cell with the nucleic acids and/or vectors.

[0021] According to the present invention, the compounds, compositions, and methods described herein can be used to increase or decrease the relative percentages of a lipid in a seed oil, increase or decrease the level of a lipid in a seed oil, or to increase or decrease the level of a fatty acid in a seed oil, or to increase or decrease the level of a starch or other carbohydrate in a seed or plant, or to increase or decrease the level of proteins in a seed or plant. The manipulations described herein can also be used to improve seed germination and growth of the young seedlings and plants and to enhance plant yield of seed storage compounds.

[0022] A method of producing a higher or lower than normal or typical level of storage compound in a transgenic plant comprises expressing a LMP nucleic acid from *Physcomitrella patens*, *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* in the transgenic plant, wherein the transgenic plant is *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* or a species different from *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa*. Also included herein are compositions and methods of the modification of the efficiency of production of a seed storage compound. As

used herein, where the phrase *Physcomitrella patens*, *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* is used, this also means *Physcomitrella patens* and/or *Arabidopsis thaliana* and/or *Brassica napus* and/or *Glycine max* and/or *Zea mays* and/or *Oryza sativa*.

[0023] Accordingly, it is an object of the present invention to provide novel isolated LMP nucleic acids and isolated LMP amino acid sequences from *Physcomitrella patens*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* as well as active fragments, analogs, and orthologs thereof. Those active fragments, analogs, and orthologs can also be from different plant species, as one skilled in the art will appreciate that other plant species will also contain those or related nucleic acids.

[0024] It is another object of the present invention to provide transgenic plants having modified levels of seed storage compounds, and in particular, modified levels of a lipid, a fatty acid, or a sugar.

[0025] The polynucleotides and polypeptides of the present invention, including agonists and/or fragments thereof, have also uses that include modulating plant growth, and potentially plant yield, preferably increasing plant growth under adverse conditions (drought, cold, light, UV). In addition, antagonists of the present invention may have uses that include modulating plant growth and/or yield, through preferably increasing plant growth and yield. In yet another embodiment, over-expression polypeptides of the present invention using a constitutive promoter (e.g., 35S, or other promoters) may be useful for increasing plant yield under stress conditions (drought, light, cold, UV) by modulating light utilization efficiency.

[0026] It is a further object of the present invention to provide methods for producing such aforementioned transgenic plants.

[0027] It is another object of the present invention to provide seeds and seed oils from such aforementioned transgenic plants.

[0028] These and other objects, features and advantages of the present invention will become apparent after a review of the following detailed description of the disclosed embodiments and the appended claims.

### DETAILED DESCRIPTION OF THE INVENTION

[0029] The present invention may be understood more readily by reference to the following detailed description of the preferred embodiments of the invention and the Examples included therein.

[0030] Before the present compounds, compositions, and methods are disclosed and described, it is to be understood that this invention is not limited to specific nucleic acids, specific polypeptides, specific cell types, specific host cells, specific conditions, or specific methods, etc., as such may, of course, vary, and the numerous modifications and variations therein will be apparent to those skilled in the art. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting. As used in the specification and in the claims, "a" or "an" can mean one or more, depending upon the context in which it is used. Thus, for example, reference to "a cell" can mean that at least one cell can be utilized.

[0031] In accordance with the purpose(s) of this invention, as embodied and broadly described herein, this invention, in one aspect, provides an isolated nucleic acid from a moss or plant (*Physcomitrella patens*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa*) encoding a Lipid Metabolism Protein (LMP), or a portion thereof.

[0032] The present invention is based, in part, on the isolation and characterization of nucleic acid molecules encoding casein kinase I-like polypeptides from a moss (*Physcomitrella patens*) or plants including canola (*Brassica napus*), soybean (*Glycine max*), corn (*Zea mays*), and rice (*Oryza sativa*).

[0033] One aspect of the invention pertains to isolated nucleic acid molecules that encode LMP polypeptides or biologically active portions thereof, and nucleic acid fragments sufficient for use as hybridization probes or primers for the identification or amplification of an LMP-encoding nucleic acid (e.g., LMP DNA). As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. This term also encompasses untranslated sequence located at both the 3' and 5' ends of the coding region of a gene: at least about 1000 nucleotides of sequence upstream from the 5' end of the coding region and at least about 200 nucleotides of sequence downstream from the 3' end of the coding region of the gene. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA. An "isolated" nucleic acid molecule is one that is substantially separated from other nucleic acid molecules, which are present in the natural source of the

nucleic acid. Preferably, an "isolated" nucleic acid is substantially free of sequences, which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism, from which the nucleic acid is derived. For example, in various embodiments, the isolated LMP nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences, which naturally flank the nucleic acid molecule in genomic DNA of the cell, from which the nucleic acid is derived (e.g., a *Physcomitrella patens*, *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays* or *Oryza sativa* cell). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized.

[0034] A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having a nucleotide sequence of Appendix A, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. For example, an *Physcomitrella patens*, *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* LMP cDNA can be isolated from an *Physcomitrella patens*, *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays* or *Oryza sativa* library using all or portion of one of the sequences of Appendix A as a hybridization probe and standard hybridization techniques (e.g., as described in Sambrook et al. 1989, *Molecular Cloning: A Laboratory Manual. 2nd, ed.*, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). Moreover, a nucleic acid molecule encompassing all or a portion of one of the sequences of Appendix A can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this sequence (e.g., a nucleic acid molecule encompassing all or a portion of one of the sequences of Appendix A can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this same sequence of Appendix A). For example, mRNA can be isolated from plant cells (e.g., by the guanidinium-thiocyanate extraction procedure of Chirgwin et al., 1979, *Biochemistry* 18:5294-5299) and cDNA can be prepared using reverse transcriptase (e.g., Moloney MLV reverse transcriptase, available from Gibco/BRL, Bethesda, MD; or AMV reverse transcriptase, available from Seikagaku America, Inc., St. Petersburg, FL). Synthetic oligonucleotide primers for polymerase chain reaction amplification can be designed based upon one of the nucleotide sequences shown in Appendix A. A nucleic acid of the invention can be amplified using cDNA or, alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector

and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to a LMP nucleotide sequence can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

[0035] In a preferred embodiment, an isolated nucleic acid of the invention comprises one of the nucleotide sequences shown in Appendix A. The sequences of Appendix A correspond to the *Physcomitrella patens*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* LMP sequences of the invention. These cDNAs comprise sequences encoding LMPs (i.e., the "coding region," indicated in Appendix A), as well as 5' untranslated sequences and 3' untranslated sequences. Alternatively, the nucleic acid molecules can comprise only the coding region of any of the sequences in Appendix A or can contain whole genomic fragments isolated from genomic DNA.

[0036] For the purposes of this application, it will be understood that each of the sequences set forth in Appendix A has an identifying entry number (e.g., BnCk01). Each of these sequences may generally comprise three parts: a 5' upstream region, a coding region, and a downstream region. A coding region of these sequences is indicated as "ORF position" (Table 3).

[0037] In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule, which is a complement of one of the nucleotide sequences shown in Appendix A, or a portion thereof. A nucleic acid molecule, which is complementary to one of the nucleotide sequences shown in Appendix A, is one which is sufficiently complementary to one of the nucleotide sequences shown in Appendix A such that it can hybridize to one of the nucleotide sequences shown in Appendix A, thereby forming a stable duplex.

[0038] In still another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which is at least about 50-60%, preferably at least about 60-70%, more preferably at least about 70-80%, 80-90%, or 90-95%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to a nucleotide sequence shown in Appendix A, or a portion thereof. In an additional preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to one of the nucleotide sequences shown in Appendix A, or a portion thereof. These hybridization conditions include washing with a solution having a salt concentration of about 0.02 molar at pH 7 at about 60°C.

[0039] Moreover, the nucleic acid molecule of the invention can comprise only a portion of the coding region of one of the sequences in Appendix A, for example a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of a LMP. The nucleotide sequences determined from the cloning of the LMP genes from *Physcomitrella patens*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* allows for the generation of probes and primers designed for use in identifying and/or cloning LMP homologues in other cell types and organisms, as well as LMP homologues from other plants or related species. Therefore, this invention also provides compounds comprising the nucleic acids disclosed herein, or fragments thereof. These compounds include the nucleic acids attached to a moiety. These moieties include, but are not limited to, detection moieties, hybridization moieties, purification moieties, delivery moieties, reaction moieties, binding moieties, and the like. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 40, 50, or 75 consecutive nucleotides of a sense strand of one of the sequences set forth in Appendix A, an anti-sense sequence of one of the sequences set forth in Appendix A, or naturally occurring mutants thereof. Primers based on a nucleotide sequence of Appendix A can be used in PCR reactions to clone LMP homologues. Probes based on the LMP nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, e.g. the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a genomic marker test kit for identifying cells which express a LMP, such as by measuring a level of a LMP-encoding nucleic acid in a sample of cells, e.g., detecting LMP mRNA levels, or determining whether a genomic LMP gene has been mutated or deleted.

[0040] In one embodiment, the nucleic acid molecule of the invention encodes a protein or portion thereof, which includes an amino acid sequence, which is sufficiently homologous to an amino acid encoded by a sequence of Appendix A such that the protein or portion thereof maintains the same or a similar function as the wild-type protein. As used herein, the language "sufficiently homologous" refers to proteins or portions thereof, which have amino acid sequences, which include a minimum number of identical or equivalent (e.g., an amino acid residue, which has a similar side chain as an amino acid residue in one of the ORFs of a sequence of Appendix A) amino acid residues to an amino acid sequence such that the protein or portion thereof is able to participate in the metabolism of compounds necessary for the

production of seed storage compounds in plants, construction of cellular membranes in microorganisms or plants, or in the transport of molecules across these membranes. Regulatory proteins, such as DNA binding proteins, transcription factors, kinases, phosphatases, or protein members of metabolic pathways such as the lipid, starch, and protein biosynthetic pathways, or membrane transport systems, may play a role in the biosynthesis of seed storage compounds. Examples of such activities are described herein (see putative annotations in Table 3). Examples of LMP-encoding nucleic acid sequences are set forth in Appendix A.

**[0041]** As altered or increased sugar and/or fatty acid production is a general trait wished to be inherited into a wide variety of plants like maize, wheat, rye, oat, triticale, rice, barley, soybean, peanut, cotton, canola, manihot, pepper, sunflower and tagetes, solanaceous plants like potato, tobacco, eggplant, and tomato, vicia species, pea, alfalfa, bushy plants (coffee, cacao, tea), Salix species, trees (oil palm, coconut), and perennial grasses and forage crops, these crop plants are also preferred target plants for genetic engineering as one further embodiment of the present invention.

**[0042]** Portions of proteins encoded by the LMP nucleic acid molecules of the invention are preferably biologically active portions of one of the LMPs. As used herein, the term "biologically active portion of a LMP" is intended to include a portion, e.g., a domain/motif, of a LMP that participates in the metabolism of compounds necessary for the biosynthesis of seed storage lipids, or the construction of cellular membranes in microorganisms or plants, or in the transport of molecules across these membranes, or has an activity as set forth in Table 3. To determine whether a LMP or a biologically active portion thereof can participate in the metabolism of compounds necessary for the production of seed storage compounds and cellular membranes, an assay of enzymatic activity may be performed. Such assay methods are well known to those skilled in the art, and as described in Example 14 of the Exemplification.

**[0043]** Biologically active portions of a LMP include peptides comprising amino acid sequences derived from the amino acid sequence of a LMP (e.g., an amino acid sequence encoded by a nucleic acid of Appendix A or the amino acid sequence of a protein homologous to a LMP, which include fewer amino acids than a full length LMP or the full length protein, which is homologous to a LMP) and exhibit at least one activity of a LMP. Typically, biologically active portions (peptides, e.g., peptides which are, for example, 5, 10, 15, 20, 30, 35, 36, 37, 38, 39, 40, 50, 100, or more amino acids in length) comprise a domain or motif with at least one activity of a LMP. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of

the activities described herein. Preferably, the biologically active portions of a LMP include one or more selected domains/motifs or portions thereof having biological activity.

[0044] Additional nucleic acid fragments encoding biologically active portions of a LMP can be prepared by isolating a portion of one of the sequences, expressing the encoded portion of the LMP or peptide (e.g., by recombinant expression *in vitro*), and assessing the activity of the encoded portion of the LMP or peptide.

[0045] The invention further encompasses nucleic acid molecules that differ from one of the nucleotide sequences shown in Appendix A (and portions thereof) due to degeneracy of the genetic code and thus encode the same LMP as that encoded by the nucleotide sequences shown in Appendix A. In a further embodiment, the nucleic acid molecule of the invention encodes a full length protein, which is substantially homologous to an amino acid sequence of a polypeptide encoded by an open reading frame shown in Appendix A. In one embodiment, the full-length nucleic acid or protein or fragment of the nucleic acid or protein is from *Physcomitrella patens*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa*.

[0046] In addition to the *Physcomitrella patens*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* LMP nucleotide sequences shown in Appendix A, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of LMPs may exist within a population (e.g., the *Physcomitrella patens*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* population). Such genetic polymorphism in the LMP gene may exist among individuals within a population due to natural variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a LMP, preferably a *Physcomitrella patens*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* LMP. Such natural variations can typically result in 1-40% variance in the nucleotide sequence of the LMP gene. Any and all such nucleotide variations, and resulting amino acid polymorphisms in LMP that are the result of natural variation and that do not alter the functional activity of LMPs, are intended to be within the scope of the invention.

[0047] Nucleic acid molecules corresponding to natural variants and non-*Physcomitrella patens*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* orthologs of the *Physcomitrella patens*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* LMP cDNA of the invention can be isolated based on their homology to *Physcomitrella patens*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* LMP nucleic acid disclosed herein using the *Physcomitrella patens*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* cDNA, or a portion thereof, as a

hybridization probe according to standard hybridization techniques under stringent hybridization conditions. As used herein, the term "orthologs" refers to two nucleic acids from different species, but that have evolved from a common ancestral gene by speciation. Normally, orthologs encode proteins having the same or similar functions. Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising a nucleotide sequence of Appendix A. In other embodiments, the nucleic acid is at least 30, 50, 100, 250 or more nucleotides in length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing, under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 65%, more preferably at least about 70%, and even more preferably at least about 75% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y., 1989: 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a sequence of Appendix A corresponds to a naturally occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein). In one embodiment, the nucleic acid encodes a natural *Physcomitrella patens*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* LMP.

**[0048]** In addition to naturally-occurring variants of the LMP sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into a nucleotide sequence of Appendix A, thereby leading to changes in the amino acid sequence of the encoded LMP without altering the functional ability of the LMP. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in a sequence of Appendix A. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of one of the LMPs (Appendix A) without altering the activity of said LMP, whereas an "essential" amino acid residue is required for LMP activity. Other amino acid residues, however, (e.g., those that are not conserved or only semi-conserved in the domain having LMP activity) may not be essential for activity, and thus are likely to be amenable to alteration without altering LMP activity.

[0049] Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding LMPs that contain changes in amino acid residues that are not essential for LMP activity. Such LMPs differ in amino acid sequence from a sequence yet retain at least one of the LMP activities described herein. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 50% homologous to an amino acid sequence encoded by a nucleic acid of Appendix A and is capable of participation in the metabolism of compounds necessary for the production of seed storage compounds in *Physcomitrella patens*, *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa*, or cellular membranes, or has one or more activities set forth in Table 3. Preferably, the protein encoded by the nucleic acid molecule is at least about 50-60% homologous to one of the sequences encoded by a nucleic acid of Appendix A, more preferably at least about 60-70% homologous to one of the sequences encoded by a nucleic acid of Appendix A, even more preferably at least about 70-80%, 80-90%, 90-95% homologous to one of the sequences encoded by a nucleic acid of Appendix A, and most preferably at least about 96%, 97%, 98%, or 99% homologous to one of the sequences encoded by a nucleic acid of Appendix A.

[0050] To determine the percent homology of two amino acid sequences (e.g., one of the sequences encoded by a nucleic acid of Appendix A and a mutant form thereof) or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of one protein or nucleic acid for optimal alignment with the other protein or nucleic acid). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in one sequence (e.g., one of the sequences encoded by a nucleic acid of Appendix A) is occupied by the same amino acid residue or nucleotide as the corresponding position in the other sequence (e.g., a mutant form of the sequence selected from the polypeptide encoded by a nucleic acid of Appendix A), then the molecules are homologous at that position (i.e., as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity"). The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % homology = numbers of identical positions/total numbers of positions x 100).

[0051] An isolated nucleic acid molecule encoding a LMP homologous to a protein sequence encoded by a nucleic acid of Appendix A can be created by introducing one or more nucleotide substitutions, additions, or deletions into a nucleotide sequence of Appendix A such that one or more amino acid substitutions, additions, or deletions are introduced into the encoded

protein. Mutations can be introduced into one of the sequences of Appendix A by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine), and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in a LMP is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a LMP coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for a LMP activity described herein to identify mutants that retain LMP activity. Following mutagenesis of one of the sequences of Appendix A, the encoded protein can be expressed recombinantly, and the activity of the protein can be determined using, for example, assays described herein (see Examples 12-14 of the Exemplification).

**[0052]** LMPs are preferably produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the protein is cloned into an expression vector (as described above), the expression vector is introduced into a host cell (as described herein), and the LMP is expressed in the host cell. The LMP can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Alternative to recombinant expression, a LMP or peptide thereof can be synthesized chemically using standard peptide synthesis techniques. Moreover, native LMP can be isolated from cells, for example using an anti-LMP antibody, which can be produced by standard techniques utilizing a LMP or fragment thereof of this invention.

**[0053]** The invention also provides LMP chimeric or fusion proteins. As used herein, a LMP "chimeric protein" or "fusion protein" comprises a LMP polypeptide operatively linked to a non-LMP polypeptide. An "LMP polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a LMP, whereas a "non-LMP polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous

to the LMP, e.g., a protein which is different from the LMP and which is derived from the same or a different organism. Within the fusion protein, the term "operatively linked" is intended to indicate that the LMP polypeptide and the non-LMP polypeptide are fused to each other so that both sequences fulfill the proposed function attributed to the sequence used. The non-LMP polypeptide can be fused to the N-terminus or C-terminus of the LMP polypeptide. For example, in one embodiment, the fusion protein is a GST-LMP (glutathione S-transferase) fusion protein, in which the LMP sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant LMPs. In another embodiment, the fusion protein is a LMP containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a LMP can be increased through use of a heterologous signal sequence.

[0054] Preferably, a LMP chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers, which give rise to complementary overhangs between two consecutive gene fragments, which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel et al., John Wiley & Sons:1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). An LMP-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the LMP.

[0055] In addition to the nucleic acid molecules encoding LMPs described above, another aspect of the invention pertains to isolated nucleic acid molecules, which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence, which is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire LMP coding strand, or to only a portion thereof. In one

embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding a LMP. The term "coding region" refers to the region of the nucleotide sequence comprising codons, which are translated into amino acid residues (e.g., the entire coding region of BnCh01 comprises nucleotides 1 to 996). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding LMP. The term "noncoding region" refers to 5' and 3' sequences, which flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

[0056] Given the coding strand sequences encoding LMP disclosed herein (e.g., the sequences set forth in Appendix A), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of LMP mRNA, but more preferably is an oligonucleotide, which is antisense to only a portion of the coding or noncoding region of LMP mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of LMP mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45, or 50 nucleotides in length. An antisense or sense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides, which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylamino-methyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydro-uracil, beta-D-galactosylqueosine, inosine, N-6-isopentenyladenine, 1-methyl-guanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methyl-cytosine, N-6-adenine, 7-methylguanine, 5-methyl-aminomethyluracil, 5-methoxyamino-methyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyl-uracil, 5-methoxyuracil, 2-methylthio-N-6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-

carboxypropyl) uracil, (acp3)w, and 2,6-diamino-purine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector, into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

**[0057]** In another variation of the antisense technology, a double-strand interfering RNA construct can be used to cause a down-regulation of the LMP mRNA level and LMP activity in transgenic plants. This requires transforming the plants with a chimeric construct containing a portion of the LMP sequence in the sense orientation fused to the antisense sequence of the same portion of the LMP sequence. A DNA linker region of variable length can be used to separate the sense and antisense fragments of LMP sequences in the construct.

**[0058]** The antisense nucleic acid molecules of the invention are typically administered to a cell or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a LMP to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule, which binds to DNA duplexes, through specific interactions in the major groove of the double helix. The antisense molecule can be modified such that it specifically binds to a receptor or an antigen expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecule to a peptide or an antibody which binds to a cell surface receptor or antigen. The antisense nucleic acid molecule can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs, in which the antisense nucleic acid molecule is placed under the control of a strong prokaryotic, viral, or eukaryotic including plant promoters, are preferred.

**[0059]** In yet another embodiment, the antisense nucleic acid molecule of the invention is an -anomeric nucleic acid molecule. An anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA, in which, contrary to the usual units, the strands run parallel to each other (Gaultier et al., 1987, Nucleic Acids Res. 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methyl-ribonucleotide (Inoue et al., 1987, Nucleic Acids Res. 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al., 1987, FEBS Lett. 215:327-330).

**[0060]** In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity, which are

capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff & Gerlach, 1988, Nature 334:585-591)) can be used to catalytically cleave LMP mRNA transcripts to thereby inhibit translation of LMP mRNA. A ribozyme having specificity for an LMP-encoding nucleic acid can be designed based upon the nucleotide sequence of a LMP cDNA disclosed herein (i.e., BnCh01 in Appendix A) or on the basis of a heterologous sequence to be isolated according to methods taught in this invention. For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed, in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a LMP-encoding mRNA (see, e.g., Cech et al., U.S. Patent No. 4,987,071 and Cech et al., U.S. Patent No. 5,116,742). Alternatively, LMP mRNA can be used to select a catalytic RNA<sup>a</sup> having a specific ribonuclease activity from a pool of RNA molecules (see, e.g., Bartel, D. & Szostak J.W., 1993, Science 261:1411-1418).

**[0061]** Alternatively, LMP gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of a LMP nucleotide sequence (e.g., a LMP promoter and/or enhancers) to form triple helical structures that prevent transcription of a LMP gene in target cells (See generally, Helene C., 1991, Anticancer Drug Des. 6:569-84; Helene C. et al., 1992, Ann. N.Y. Acad. Sci. 660:27-36; and Maher, L.J., 1992, Bioassays 14:807-15).

**[0062]** Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a LMP (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting or causing to be transported another nucleic acid, to which it has been linked. One type of vector is a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used inter-changeably as the plasmid is the most commonly used form of vector. However, the invention is intended to

include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

[0063] The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence and both sequences are fused to each other so that each fulfills its proposed function (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990) or see: Gruber and Crosby, in: Methods in Plant Molecular Biology and Biotechnology, CRC Press, Boca Raton, Florida, eds.: Glick & Thompson, Chapter 7, 89-108 including the references therein. Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell, and those which direct expression of the nucleotide sequence only in certain host cells or under certain conditions. It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., LMPs, mutant forms of LMPs, fusion proteins, etc.).

[0064] The recombinant expression vectors of the invention can be designed for expression of LMPs in prokaryotic or eukaryotic cells. For example, LMP genes can be expressed in bacterial cells, insect cells (using baculovirus expression vectors), yeast and other fungal cells (see Romanos M.A. et al., 1992, Foreign gene expression in yeast: a review, *Yeast* 8:423-488; van den Hondel, C.A.M.J.J. et al., 1991, Heterologous gene expression in filamentous fungi, in: *More Gene Manipulations in Fungi*, Bennet & Lasure, eds., p. 396-428, Academic Press, San Diego; and van den Hondel & Punt, 1991, Gene transfer systems and vector development for filamentous fungi, in: *Applied Molecular Genetics of Fungi*, Peberdy et al., eds., p. 1-28, Cambridge University Press: Cambridge), algae (Falciatore et al., 1999, *Marine*

Biotechnology 1:239-251), ciliates of the types: Holotrichia, Peritrichia, Spirotrichia, Suctoria, Tetrahymena, Paramecium, Colpidium, Glaucocystis, Platyophrya, Potomacus, Pseudocohnilembus, Euplotes, Engelmanniella, and Stylonychia, especially of the genus Stylonychia lemnae with vectors following a transformation method as described in WO 98/01572 and multicellular plant cells (see Schmidt & Willmitzer, 1988, High efficiency *Agrobacterium tumefaciens*-mediated transformation of *Arabidopsis thaliana* leaf and cotyledon plants, Plant Cell Rep.:583-586); Plant Molecular Biology and Biotechnology, C Press, Boca Raton, Florida, chapter 6/7, S.71-119 (1993); White, Jenes et al., Techniques for Gene Transfer, in: Transgenic Plants, Vol. 1, Engineering and Utilization, eds.: Kung and Wu, Academic Press 1993, 128-43; Potrykus, 1991, Annu. Rev. Plant Physiol. Plant Mol. Biol. 42:205-225 (and references cited therein) or mammalian cells. Suitable host cells are discussed further in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA, 1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

**[0065]** Expression of proteins in prokaryotes is most often carried out with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein but also to the C-terminus or fused within suitable regions in the proteins. Such fusion vectors typically serve one or more of the following purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase.

**[0066]** Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith & Johnson, 1988, Gene 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ), which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. In one embodiment, the coding sequence of the LMP is cloned into a pGEX expression vector to create a vector encoding a fusion protein comprising, from the N-terminus to the C-terminus, GST-thrombin cleavage site-X protein. The fusion protein can be purified by affinity chromatography using glutathione-

agarose resin. Recombinant LMP unfused to GST can be recovered by cleavage of the fusion protein with thrombin.

[0067] Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., 1988, Gene 69:301-315) and pET 11d (Studier et al., 1990, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, California 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

[0068] One strategy to maximize recombinant protein expression is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman S., 1990, Gene Expression Technology: *Methods in Enzymology* 185:119-128, Academic Press, San Diego, California). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in the bacterium chosen for expression (Wada et al., 1992, Nucleic Acids Res. 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

[0069] In another embodiment, the LMP expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari et al., 1987, Embo J. 6:229-234), pMFa (Kurjan & Herskowitz, 1982, Cell 30:933-943), pJRY88 (Schultz et al., 1987, Gene 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, CA). Vectors and methods for the construction of vectors appropriate for use in other fungi, such as the filamentous fungi, include those detailed in: van den Hondel & Punt, 1991, "Gene transfer systems and vector development for filamentous fungi, in: Applied Molecular Genetics of Fungi, Peberdy et al., eds., p. 1-28, Cambridge University Press: Cambridge.

[0070] Alternatively, the LMPs of the invention can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al., 1983, Mol. Cell Biol. 3:2156-2165) and the pVL series (Lucklow & Summers, 1989, Virology 170:31-39).

[0071] In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression

vectors include pCDM8 (Seed, 1987, Nature 329:840) and pMT2PC (Kaufman et al., 1987, EMBO J. 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, Fritsh and Maniatis, *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

[0072] In another embodiment, the LMPs of the invention may be expressed in unicellular plant cells (such as algae, see Falciatore et al., 1999, Marine Biotechnology 1:239-251, and references therein) and plant cells from higher plants (e.g., the spermatophytes, such as crop plants). Examples of plant expression vectors include those detailed in: Becker, Kemper, Schell and Masterson (1992, "New plant binary vectors with selectable markers located proximal to the left border," Plant Mol. Biol. 20:1195-1197) and Bevan (1984, "Binary *Agrobacterium* vectors for plant transformation," Nucleic Acids Res. 12:8711-8721; Vectors for Gene Transfer in Higher Plants; in: Transgenic Plants, Vol. 1, Engineering and Utilization, eds.: Kung und R. Wu, Academic Press, 1993, S. 15-38).

[0073] A plant expression cassette preferably contains regulatory sequences capable to drive gene expression in plant cells and which are operably linked so that each sequence can fulfill its function such as termination of transcription, including polyadenylation signals. Preferred polyadenylation signals are those originating from *Agrobacterium tumefaciens* t-DNA such as the gene 3 known as octopine synthase of the Ti-plasmid pTiACH5 (Gielen et al., 1984, EMBO J. 3:835) or functional equivalents thereof but also all other terminators functionally active in plants are suitable.

[0074] As plant gene expression is very often not limited on transcriptional levels a plant expression cassette preferably contains other operably linked sequences like translational enhancers such as the overdrive-sequence containing the 5'-untranslated leader sequence from tobacco mosaic virus enhancing the protein per RNA ratio (Gallie et al., 1987, Nucleic Acids Res. 15:8693-8711).

[0075] Plant gene expression has to be operably linked to an appropriate promoter conferring gene expression in a timely, cell or tissue specific manner. Preferred are promoters driving constitutive expression (Benfey et al., 1989, EMBO J. 8:2195-2202) like those derived from plant viruses like the 35S CaMV (Franck et al., 1980, Cell 21:285-294), the 19S CaMV (see also US 5,352,605 and WO 84/02913) or plant promoters like those from Rubisco small

subunit described in US 4,962,028. Even more preferred are seed-specific promoters driving expression of LMP proteins during all or selected stages of seed development. Seed-specific plant promoters are known to those of ordinary skill in the art and are identified and characterized using seed-specific mRNA libraries and expression profiling techniques. Seed-specific promoters include the napin-gene promoter from rapeseed (US 5,608,152), the USP-promoter from *Vicia faba* (Baeumlein et al., 1991, Mol. Gen. Genetics 225:459-67), the oleosin-promoter from *Arabidopsis* (WO 98/45461), the phaseolin-promoter from *Phaseolus vulgaris* (US 5,504,200), the Bce4-promoter from *Brassica* (WO9113980), or the legumin B4 promoter (LeB4; Baeumlein et al., 1992, Plant J. 2:233-239), as well as promoters conferring seed specific expression in monocot plants like maize, barley, wheat, rye, rice, etc. Suitable promoters to note are the lpt2 or lpt1-gene promoter from barley (WO 95/15389 and WO 95/23230) or those described in WO 99/16890 (promoters from the barley hordein-gene, the rice glutelin gene, the rice oryzin gene, the rice prolamin gene, the wheat gliadin gene, wheat glutelin gene, the maize zein gene, the oat glutelin gene, the Sorghum kasirin-gene, and the rye secalin gene).

[0076] Plant gene expression can also be facilitated via an inducible promoter (for review see Gatz 1997, Annu. Rev. Plant Physiol. Plant Mol. Biol. 48:89-108). Chemically inducible promoters are especially suitable if gene expression is desired in a time specific manner. Examples for such promoters are a salicylic acid inducible promoter (WO 95/19443), a tetracycline inducible promoter (Gatz et al., 1992, Plant J. 2:397-404), and an ethanol inducible promoter (WO 93/21334).

[0077] Promoters responding to biotic or abiotic stress conditions are also suitable promoters such as the pathogen inducible PRP1-gene promoter (Ward et al., 1993, Plant. Mol. Biol. 22:361-366), the heat inducible hsp80-promoter from tomato (US 5,187,267), cold inducible alpha-amylase promoter from potato (WO 96/12814) or the wound-inducible pinII-promoter (EP 375091).

[0078] Other preferred sequences for use in plant gene expression cassettes are targeting-sequences necessary to direct the gene-product in its appropriate cell compartment (for review see Kermode, 1996, Crit. Rev. Plant Sci. 15:285-423 and references cited therein) such as the vacuole, the nucleus, all types of plastids like amyloplasts, chloroplasts, chromoplasts, the extracellular space, mitochondria, the endoplasmic reticulum, oil bodies, peroxisomes, and other compartments of plant cells. Also especially suited are promoters that confer plastid-specific gene expression, as plastids are the compartment where precursors and some end products of lipid biosynthesis are synthesized. Suitable promoters such as the viral RNA-polymerase

promoter are described in WO 95/16783 and WO 97/06250 and the clpP-promoter from *Arabidopsis* described in WO 99/46394.

**[0079]** The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule, which is antisense to LMP mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity, of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub et al., 1986, *Antisense RNA as a molecular tool for genetic analysis*, Reviews - Trends in Genetics, Vol. 1, and Mol et al., 1990, *FEBS Lett.* 268:427-430.

**[0080]** Another aspect of the invention pertains to host cells, into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is to be understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein. A host cell can be any prokaryotic or eukaryotic cell. For example, a LMP can be expressed in bacterial cells, insect cells, fungal cells, mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells), algae, ciliates or plant cells. Other suitable host cells are known to those skilled in the art.

**[0081]** Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation," "transfection," "conjugation," and "transduction" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, natural competence, chemical-mediated transfer, or electroporation. Suitable methods for transforming or transfecting host cells including plant cells can be found in

Sambrook et al., 1989, *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, and other laboratory manuals such as *Methods in Molecular Biology* 1995, Vol. 44, *Agrobacterium* protocols, ed: Gartland and Davey, Humana Press, Totowa, New Jersey.

[0082] For stable transfection of mammalian and plant cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those that confer resistance to drugs, such as G418, hygromycin, kanamycin, and methotrexate, or in plants that confer resistance towards an herbicide such as glyphosate or glufosinate. A nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a LMP or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by, for example, drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

[0083] To create a homologous recombinant microorganism, a vector is prepared which contains at least a portion of a LMP gene, into which a deletion, addition, or substitution has been introduced to thereby alter, e.g., functionally disrupt, the LMP gene. Preferably, this LMP gene is an *Physcomitrella patens*, *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* LMP gene, but it can be a homologue from a related plant or even from a mammalian, yeast, or insect source. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous LMP gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a knock-out vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous LMP gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous LMP). To create a point mutation via homologous recombination, DNA-RNA hybrids can be used in a technique known as chimeraplasty (Cole-Strauss et al. 1999, *Nucleic Acids Res.* 27:1323-1330 and Kmiec 1999, *American Scientist* 87:240-247). Homologous recombination procedures in *Arabidopsis thaliana* or other crops are also well known in the art and are contemplated for use herein.

[0084] In a homologous recombination vector, the altered portion of the LMP gene is flanked at its 5' and 3' ends by additional nucleic acid of the LMP gene to allow for homologous recombination to occur between the exogenous LMP gene carried by the vector and an

endogenous LMP gene in a microorganism or plant. The additional flanking LMP nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several hundreds of base pairs up to kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see e.g., Thomas & Capecchi 1987, Cell 51:503, for a description of homologous recombination vectors). The vector is introduced into a microorganism or plant cell (e.g., via polyethyleneglycol mediated DNA). Cells in which the introduced LMP gene has homologously recombined with the endogenous LMP gene are selected using art-known techniques.

**[0085]** In another embodiment, recombinant microorganisms can be produced which contain selected systems, which allow for regulated expression of the introduced gene. For example, inclusion of a LMP gene on a vector placing it under control of the lac operon permits expression of the LMP gene only in the presence of IPTG. Such regulatory systems are well known in the art.

**[0086]** A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture can be used to produce (i.e., express) a LMP. Accordingly, the invention further provides methods for producing LMPs using the host cells of the invention. In one embodiment, the method comprises culturing a host cell of the invention (into which a recombinant expression vector encoding a LMP has been introduced, or which contains a wild-type or altered LMP gene in its genome) in a suitable medium until LMP is produced. In another embodiment, the method further comprises isolating LMPs from the medium or the host cell.

**[0087]** Another aspect of the invention pertains to isolated LMPs, and biologically active portions thereof. An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of LMP, in which the protein is separated from cellular components of the cells in which it is naturally or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of LMP having less than about 30% (by dry weight) of non-LMP (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-LMP, still more preferably less than about 10% of non-LMP, and most preferably less than about 5% non-LMP. When the LMP or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the

volume of the protein preparation. The language "substantially free of chemical precursors or other chemicals" includes preparations of LMP in which the protein is separated from chemical precursors or other chemicals, which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of LMP having less than about 30% (by dry weight) of chemical precursors or non-LMP chemicals, more preferably less than about 20% chemical precursors or non-LMP chemicals, still more preferably less than about 10% chemical precursors or non-LMP chemicals, and most preferably less than about 5% chemical precursors or non-LMP chemicals. In preferred embodiments, isolated proteins or biologically active portions thereof lack contaminating proteins from the same organism, from which the LMP is derived. Typically, such proteins are produced by recombinant expression of, for example, an *Physcomitrella patens*, *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* LMP in other plants than *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* or microorganisms, algae or fungi.

[0088] An isolated LMP or a portion thereof of the invention can participate in the metabolism of compounds necessary for the production of seed storage compounds in *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* or of cellular membranes, or has one or more of the activities set forth in Table 3. In preferred embodiments, the protein or portion thereof comprises an amino acid sequence which is sufficiently homologous to an amino acid sequence encoded by a nucleic acid of Appendix A such that the protein or portion thereof maintains the ability to participate in the metabolism of compounds necessary for the construction of cellular membranes in *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa*, or in the transport of molecules across these membranes. The portion of the protein is preferably a biologically active portion as described herein. In another preferred embodiment, a LMP of the invention has an amino acid sequence encoded by a nucleic acid of Appendix A. In yet another preferred embodiment, the LMP has an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to a nucleotide sequence of Appendix A. In still another preferred embodiment, the LMP has an amino acid sequence, which is encoded by a nucleotide sequence that is at least about 50-60%, preferably at least about 60-70%, more preferably at least about 70-80%, 80-90%, 90-95%, and even more preferably at least about 96%, 97%, 98%, 99% or more homologous to one of the amino acid sequences encoded by a nucleic acid of Appendix A. The preferred LMPs of the present invention also preferably possess at least one of the LMP

activities described herein. For example, a preferred LMP of the present invention includes an amino acid sequence encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to a nucleotide sequence of Appendix A, and which can participate in the metabolism of compounds necessary for the construction of cellular membranes in *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa*, or in the transport of molecules across these membranes, or which has one or more of the activities set forth in Table 3.

[0089] In other embodiments, the LMP is substantially homologous to an amino acid sequence encoded by a nucleic acid of Appendix A and retains the functional activity of the protein of one of the sequences encoded by a nucleic acid of Appendix A yet differs in amino acid sequence due to natural variation or mutagenesis, as described in detail above. Accordingly, in another embodiment, the LMP is a protein which comprises an amino acid sequence, which is at least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80, 80-90, 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to an entire amino acid sequence and which has at least one of the LMP activities described herein. In another embodiment, the invention pertains to a *Physcomitrella patens*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* protein, which is substantially homologous to an entire amino acid sequence encoded by a nucleic acid of Appendix A.

[0090] Dominant negative mutations or trans-dominant suppression can be used to reduce the activity of a LMP in transgenics seeds in order to change the levels of seed storage compounds. To achieve this a mutation that abolishes the activity of the LMP is created and the inactive non-functional LMP gene is overexpressed in the transgenic plant. The inactive trans-dominant LMP protein competes with the active endogenous LMP protein for substrate or interactions with other proteins and dilutes out the activity of the active LMP. In this way the biological activity of the LMP is reduced without actually modifying the expression of the endogenous LMP gene. This strategy was used by Pontier et al to modulate the activity of plant transcription factors (Pontier D, Miao ZH, Lam E, Plant J., 2001 Sep. 27(6): 529-38, Trans-dominant suppression of plant TGA factors reveals their negative and positive roles in plant defense responses).

[0091] Homologues of the LMP can be generated by mutagenesis, e.g., discrete point mutation or truncation of the LMP. As used herein, the term "homologue" refers to a variant form of the LMP, which acts as an agonist or antagonist of the activity of the LMP. An agonist of the LMP can retain substantially the same, or a subset, of the biological activities of the LMP.

An antagonist of the LMP can inhibit one or more of the activities of the naturally occurring form of the LMP, by, for example, competitively binding to a downstream or upstream member of the cell membrane component metabolic cascade which includes the LMP, or by binding to a LMP which mediates transport of compounds across such membranes, thereby preventing translocation from taking place.

[0092] In an alternative embodiment, homologues of the LMP can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of the LMP for LMP agonist or antagonist activity. In one embodiment, a variegated library of LMP variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of LMP variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential LMP sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of LMP sequences therein. There are a variety of methods which can be used to produce libraries of potential LMP homologues from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential LMP sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang, 1983, *Tetrahedron* 39:3; Itakura et al., 1984, *Annu. Rev. Biochem.* 53:323; Itakura et al., 1984, *Science* 198:1056; Ike et al., 1983, *Nucleic Acids Res.* 11:477).

[0093] In addition, libraries of fragments of the LMP coding sequences can be used to generate a variegated population of LMP fragments for screening and subsequent selection of homologues of a LMP. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a LMP coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA, which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the LMP.

[0094] Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries

for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of LMP homologues. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique, which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify LMP homologues (Arkin & Yourvan, 1992, Proc. Natl. Acad. Sci. USA 89:7811-7815; Delgrave et al., 1993, Protein Engineering 6:327-331).

[0095] In another embodiment, cell based assays can be exploited to analyze a variegated LMP library, using methods well known in the art.

[0096] The nucleic acid molecules, proteins, protein homologues, fusion proteins, primers, vectors, and host cells described herein can be used in one or more of the following methods: identification of *Physcomitrella patens*, *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* and related organisms; mapping of genomes of organisms related to *Physcomitrella patens*, *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa*; identification and localization of *Physcomitrella patens*, *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* sequences of interest; evolutionary studies; determination of LMP regions required for function; modulation of a LMP activity; modulation of the metabolism of one or more cell functions; modulation of the transmembrane transport of one or more compounds; and modulation of seed storage compound accumulation.

[0097] The moss *Physcomitrella patens* represents one member of seed storage compounds producing organisms. It is related to plants such as *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* which require light to drive photosynthesis and growth. Plants like *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* share a high degree of homology on the DNA sequence and polypeptide level, allowing the use of heterologous screening of DNA molecules with probes evolving from other plants or organisms, thus enabling the derivation of a consensus sequence suitable for heterologous screening or functional annotation and prediction of gene functions in third species. The ability to identify such functions can therefore have significant relevance, e.g., prediction of substrate

specificity of enzymes. Further, these nucleic acid molecules may serve as reference points for the mapping of *Arabidopsis* genomes, or of genomes of related organisms.

[0098] The LMP nucleic acid molecules of the invention have a variety of uses. First, they may be used to identify an organism as being *Physcomitrella patens*, *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* or a close relative thereof. Also, they may be used to identify the presence of *Physcomitrella patens*, *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* or a relative thereof in a mixed population of microorganisms. The invention provides the nucleic acid sequences of a number of *Physcomitrella patens*, *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* genes; by probing the extracted genomic DNA of a culture of a unique or mixed population of microorganisms under stringent conditions with a probe spanning a region of a *Physcomitrella patens*, *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* gene, which is unique to this organism, one can ascertain whether this organism is present.

[0099] Further, the nucleic acid and protein molecules of the invention may serve as markers for specific regions of the genome. This has utility not only in the mapping of the genome, but also for functional studies of *Physcomitrella patens*, *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* proteins. For example, to identify the region of the genome, to which a particular *Physcomitrella patens*, *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* DNA-binding protein binds, the *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* genome could be digested, and the fragments incubated with the DNA-binding protein. Those which bind the protein may be additionally probed with the nucleic acid molecules of the invention, preferably with readily detectable labels; binding of such a nucleic acid molecule to the genome fragment enables the localization of the fragment to the genome map of *Physcomitrella patens*, *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa*, and, when performed multiple times with different enzymes, facilitates a rapid determination of the nucleic acid sequence, to which the protein binds. Further, the nucleic acid molecules of the invention may be sufficiently homologous to the sequences of related species such that these nucleic acid molecules may serve as markers for the construction of a genomic map in related plants.

[00100] The LMP nucleic acid molecules of the invention are also useful for evolutionary and protein structural studies. The metabolic and transport processes in which the molecules of the invention participate are utilized by a wide variety of prokaryotic and eukaryotic cells; by comparing the sequences of the nucleic acid molecules of the present invention to those encoding

similar enzymes from other organisms, the evolutionary relatedness of the organisms can be assessed. Similarly, such a comparison permits an assessment of which regions of the sequence are conserved and which are not, which may aid in determining those regions of the protein which are essential for the functioning of the enzyme. This type of determination is of value for protein engineering studies and may give an indication of what the protein can tolerate in terms of mutagenesis without losing function.

**[00101]** Manipulation of the LMP nucleic acid molecules of the invention may result in the production of LMPs having functional differences from the wild-type LMPs. These proteins may be improved in efficiency or activity, may be present in greater numbers in the cell than is usual, or may be decreased in efficiency or activity.

**[00102]** There are a number of mechanisms by which the alteration of a LMP of the invention may directly affect the accumulation and/or composition of seed storage compounds. In the case of plants expressing LMPs, increased transport can lead to altered accumulation of compounds and/or solute partitioning within the plant tissue and organs which ultimately could be used to affect the accumulation of one or more seed storage compounds during seed development. An example is provided by Mitsukawa et al., 1997, Proc. Natl. Acad. Sci. USA 94:7098-7102, where over expression of an Arabidopsis high-affinity phosphate transporter gene in tobacco cultured cells enhanced cell growth under phosphate-limited conditions. Phosphate availability also affects significantly the production of sugars and metabolic intermediates (Hurry et al., 2000, Plant J. 24:383-396) and the lipid composition in leaves and roots (Härtel et al., 2000, Proc. Natl. Acad. Sci. USA 97:10649-10654). Likewise, the activity of the plant ACCase has been demonstrated to be regulated by phosphorylation (Savage & Ohlrogge, 1999, Plant J. 18:521-527) and alterations in the activity of the kinases and phosphatases (LMPs) that act on the ACCase could lead to increased or decreased levels of seed lipid accumulation. Moreover, the presence of lipid kinase activities in chloroplast envelope membranes suggests that signal transduction pathways and/or membrane protein regulation occur in envelopes (see, e.g., Müller et al., 2000, J. Biol. Chem. 275:19475-19481 and literature cited therein). The *ABI1* and *ABI2* genes encode two protein serine/threonine phosphatases 2C, which are regulators in abscisic acid signaling pathway, and thereby in early and late seed development (e.g. Merlot et al., 2001, Plant J. 25:295-303). For more examples see also the section "Background of the Invention."

**[00103]** The present invention also provides antibodies, which specifically binds to an LMP-polypeptide, or a portion thereof, as encoded by a nucleic acid disclosed herein or as described herein.

**[00104]** Antibodies can be made by many well-known methods (see, e.g. Harlow and Lane, "Antibodies; A Laboratory Manual," Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1988). Briefly, purified antigen can be injected into an animal in an amount and in intervals sufficient to elicit an immune response. Antibodies can either be purified directly, or spleen cells can be obtained from the animal. The cells can then fused with an immortal cell line and screened for antibody secretion. The antibodies can be used to screen nucleic acid clone libraries for cells secreting the antigen. Those positive clones can then be sequenced (see, for example, Kelly et al., 1992, *Bio/Technology* 10:163-167; Bebbington et al., 1992, *Bio/Technology* 10:169-175).

**[00105]** The phrase "selectively binds" with the polypeptide refers to a binding reaction, which is determinative of the presence of the protein in a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bound to a particular protein do not bind in a significant amount to other proteins present in the sample. Selective binding to an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. A variety of immunoassay formats may be used to select antibodies that selectively bind with a particular protein. For example, solid-phase ELISA immuno-assays are routinely used to select antibodies selectively immunoreactive with a protein. See Harlow and Lane "Antibodies, A Laboratory Manual," Cold Spring Harbor Publications, New York (1988), for a description of immunoassay formats and conditions that could be used to determine selective binding.

**[00106]** In some instances, it is desirable to prepare monoclonal antibodies from various hosts. A description of techniques for preparing such monoclonal antibodies may be found in Stites et al., editors, "Basic and Clinical Immunology," Lange Medical Publications, Los Altos, Calif., Fourth Edition (and references cited therein), and in Harlow and Lane, "Antibodies, A Laboratory Manual," Cold Spring Harbor Publications, New York, 1988.

**[00107]** Throughout this application, various publications are referenced. The disclosures of all of these publications and those references cited within those publications in their entireties are hereby incorporated by reference into this application in order to more fully describe the state of the art, to which this invention pertains.

**[00108]** It will be apparent to those skilled in the art that various modifications and variations can be made in the present invention without departing from the scope or spirit of the invention. Other embodiments of the invention will be apparent to those skilled in the art from consideration of the specification and practice of the invention disclosed herein. It is intended

that the specification and Examples be considered as exemplary only, with a true scope and spirit of the invention being indicated by the claims included herein.

## EXAMPLES

### Example 1: General Processes

#### a) General Cloning Processes:

[00109] Cloning processes such as, for example, restriction cleavages, agarose gel electrophoresis, purification of DNA fragments, transfer of nucleic acids to nitrocellulose and nylon membranes, linkage of DNA fragments, transformation of *Escherichia coli* and yeast cells, growth of bacteria and sequence analysis of recombinant DNA were carried out as described in Sambrook et al. (1989, Cold Spring Harbor Laboratory Press: ISBN 0-87969-309-6) or Kaiser, Michaelis and Mitchell (1994, "Methods in Yeast Genetics", Cold Spring Harbor Laboratory Press: ISBN 0-87969-451-3).

#### b) Chemicals:

[00110] The chemicals used were obtained, if not mentioned otherwise in the text, in p.a. quality from the companies Fluka (Neu-Ulm), Merck (Darmstadt), Roth (Karlsruhe), Serva (Heidelberg) and Sigma (Deisenhofen). Solutions were prepared using purified, pyrogen-free water, designated as H<sub>2</sub>O in the following text, from a Milli-Q water system water purification plant (Millipore, Eschborn). Restriction endonucleases, DNA-modifying enzymes and molecular biology kits were obtained from the companies AGS (Heidelberg), Amersham (Braunschweig), Biometra (Göttingen), Boehringer (Mannheim), Genomed (Bad Oeynhausen), New England Biolabs (Schwalbach/ Taunus), Novagen (Madison, Wisconsin, USA), Perkin-Elmer (Weiterstadt), Pharmacia (Freiburg), Qiagen (Hilden) and Stratagene (Amsterdam, Netherlands). They were used, if not mentioned otherwise, according to the manufacturer's instructions.

#### c) Plant Material and Growth:

##### *Physcomitrella patens*

[00111] For this study, plants of the species *Physcomitrella patens* (Hedw.) B.S.G. from the collection of the genetic studies section of the University of Hamburg were used. They originate from the strain 16/14 collected by H.L.K. Whitehouse in Gransden Wood,

Huntingdonshire (England), which was subcultured from a spore by Engel (1968, *Am. J. Bot.* 55: 438-446). Proliferation of the plants was carried out by means of spores and by means of regeneration of the gametophytes. The protonema developed from the haploid spore as a chloroplast-rich chloronema and chloroplast-low caulonema, on which buds formed after approximately 12 days. These grew to give gametophores bearing antheridia and archegonia. After fertilization, the diploid sporophyte with a short seta and the spore capsule resulted, in which the meiospores mature.

[00112] Culturing was carried out in a climatic chamber at an air temperature of 25°C and light intensity of 55  $\mu\text{mol m}^{-2} \text{s}^{-1}$  (white light; Philips TL 65W/25 fluorescent tube) and a light/dark change of 16/8 hours.

#### *Arabidopsis thaliana*

[00113] *Arabidopsis thaliana* cv Columbia were grown on plates with half-strength MS medium (Murashige & Skoog, 1962, *Physiol. Plant.* 15, 473-497), pH 6.2, 2% sucrose and 0.8% agar. Seeds were sterilized for 20 minutes in 20% bleach 0.5% triton X100 and rinsed 6 times with excess sterile water. *Arabidopsis thaliana* seeds were preincubated for three days in the dark at 4°C before placing them into an incubator (AR-75, Percival Scientific, Boone, IA) at a photon flux density of 60-80  $\mu\text{mol m}^{-2} \text{s}^{-1}$  and a light period of 16 hours (22°C), and a dark period of 8 hours (18°C). Plants were either grown as described above or on soil under standard conditions as described in Focks & Benning (1998, *Plant Physiol.* 118:91-101).

#### *Brassica napus*

[00114] *Brassica napus* varieties AC Excel and Cresor were used for this study to create cDNA libraries. Seed, seed pod, flower, leaf, stem and root tissues were collected from plants that were in some cases dark-, salt-, heat -and drought-treated. However, this study focused on the use of seed and seed pod tissues for cDNA libraries. Plants were tagged to harvest seeds collected 60 - 75 days after planting from two time points: 1-15 days and 15 -25 days after anthesis. Plants have been grown in Metromix (Scotts, Marysville, OH) at 71°F under a 14 hr photoperiod. Six seed and seed pod tissues of interest in this study were collected to create the following cDNA libraries: Immature seeds, mature seeds, immature seed pods, mature seed pods, night-harvested seed pods and Cresor variety (high erucic acid) seeds. Tissue samples were collected within specified time points for each developing tissue and multiple samples within a time frame pooled together for eventual extraction of total RNA. Samples from

immature seeds were taken between 1-25 days after anthesis (daa), mature seeds between 25-50 daa, immature seed pods between 1-15 daa, mature seed pods between 15-50 daa, night-harvested seed pods between 1-50 daa and Cresor seeds 5-25 daa.

### *Glycine max*

[00115] *Glycine max* cv. Resnick was used for this study to create cDNA libraries. Seed, seed pod, flower, leaf, stem and root tissues were collected from plants that were in some cases dark-, salt-, heat- and drought-treated. In some cases plants have been nematode infected as well. However, this study focused on the use of seed and seed pod tissues for cDNA libraries. Plants were tagged to harvest seeds at the set days after anthesis: 5-15, 15-25, 25-35, & 33-50.

### *Zea mays*

[00116] *Zea mays* hybrid B73 x Mo17 and B73 inbred (the female inbred parent of the hybrid B73 x Mo17) were used to generate cDNA libraries. Fruit or Seed (Fertilized ovules/young kernels at stage 1 and 9 d post pollination; kernels at milk stage [R3, early starch production], 23 d post pollination; kernels at early dough stage (R4), developing starch grains and well-formed embryo present, 30 d post pollination of field-grown plants; very young kernels at blister stage [R2, watery endosperm]; kernels at early dent stage (R5), endosperm becoming firm, 36 d post pollination; B73 inbreds, kernels at 9 and 19 d post pollination), flowers (tassel development: from 6 cm tassel (V10) up to and including anthesis, 44 to 70 dap; ear development: ear shoots from 2 cm (V13) up to and including silking (unpollinated), 51 to 70 dap), leaves/shoot/rosettes (mixed ages, all prior to seed-fill; includes leaves of a) 3-leaf plants (V3), b) 6-leaf plants (V6), and c) an older source leaf (3rd from the ground), just before tassel emergence in the field), stem (located underground of 2 to 5-leaf plants; roots and most leaf tissue removed, 13 to 29 dap of field-grown plants; Stem tissue near the ear at tassel emergence and during seed-fill (milk stage), 56 to 84 dap, field-grown plants) and root tissues (from young to mid-age plants: from seedlings, 6-leaf plants, and 9-leaf plants; 12 to 35 dap) were collected from plants.

### *Oryza sativa*

[00117] *Oryza sativa* ssp. Japonica cv. Nippon-barre was used for this study to create cDNA libraries. Seed, seed pod, flower, leaf, stem and root tissues were collected from plants that were in some cases dark-, salt-, heat- and drought-treated. This study focused on the use of

seed embryo tissues for cDNA libraries. Embryo and endosperm were collected separately in case endosperm tissue might interfere with RNA extraction. Plants have been grown in the greenhouse on Wisconsin soil (has high organic matter) at 85°F under a 14-h photoperiod. Rice embryos were dissected out of the developing seeds.

#### **Example 2: Total DNA Isolation from Plants**

[00118] The details for the isolation of total DNA relate to the working up of one gram fresh weight of plant material.

[00119] CTAB buffer: 2% (w/v) N-cethyl-N,N,N-trimethylammonium bromide (CTAB); 100 mM Tris HCl pH 8.0; 1.4 M NaCl; 20 mM EDTA. N-Laurylsarcosine buffer: 10% (w/v) N-laurylsarcosine; 100 mM Tris HCl pH 8.0; 20 mM EDTA.

[00120] The plant material was triturated under liquid nitrogen in a mortar to give a fine powder and transferred to 2 ml Eppendorf vessels. The frozen plant material was then covered with a layer of 1 ml of decomposition buffer (1 ml CTAB buffer, 100 µl of N-laurylsarcosine buffer, 20 µl of β-mercaptoethanol and 10 µl of proteinase K solution, 10 mg/ml) and incubated at 60°C for one hour with continuous shaking. The homogenate obtained was distributed into two Eppendorf vessels (2 ml) and extracted twice by shaking with the same volume of chloroform/isoamyl alcohol (24:1). For phase separation, centrifugation was carried out at 8000g and RT for 15 min in each case. The DNA was then precipitated at -70°C for 30 min using ice-cold isopropanol. The precipitated DNA was sedimented at 4°C and 10,000g for 30 min and resuspended in 180 µl of TE buffer (Sambrook et al., 1989, Cold Spring Harbor Laboratory Press: ISBN 0-87969-309-6). For further purification, the DNA was treated with NaCl (1.2 M final concentration) and precipitated again at -70°C for 30 min using twice the volume of absolute ethanol. After a washing step with 70% ethanol, the DNA was dried and subsequently taken up in 50 µl of H<sub>2</sub>O + RNase (50 mg/ml final concentration). The DNA was dissolved overnight at 4°C and the RNase digestion was subsequently carried out at 37°C for 1 h. Storage of the DNA took place at 4°C.

#### **Example 3: Isolation of Total RNA and poly-(A)+ RNA from Plants**

##### ***Arabidopsis thaliana***

[00121] For the investigation of transcripts, both total RNA and poly-(A)+ RNA were isolated.

[00122] RNA is isolated from siliques of *Arabidopsis* plants according to the following procedure:

[00123] RNA preparation from *Arabidopsis* seeds - "hot" extraction:

1. Buffers, enzymes and solution

- 2M KCl
- Proteinase K
- Phenol (for RNA)
- Chloroform:Isoamylalcohol  
(Phenol:chloroform 1:1; pH adjusted for RNA)
- 4 M LiCl, DEPC-treated
- DEPC-treated water
- 3M NaOAc, pH 5, DEPC-treated
- Isopropanol
- 70% ethanol (made up with DEPC-treated water)
- Resuspension buffer:0.5% SDS, 10 mM Tris pH 7.5, 1 mM EDTA made up with DEPC-treated water as this solution can not be DEPC-treated
- Extraction Buffer:  
0.2M Na Borate  
30 mM EDTA  
30 mM EGTA  
1% SDS (250 $\mu$ l of 10% SDS-solution for 2.5ml buffer)  
1% Deoxycholate (25mg for 2,5ml buffer)  
2% PVPP (insoluble - 50mg for 2.5ml buffer)  
2% PVP 40K (50mg for 2.5ml buffer)  
10 mM DTT

100 mM  $\beta$ -Mercaptoethanol (fresh, handle under fume hood - use 35 $\mu$ l of 14.3M solution for 5ml buffer)

2. Extraction

[00124] Heat extraction buffer up to 80°C. Grind tissue in liquid nitrogen-cooled mortar, transfer tissue powder to 1.5ml tube. Tissue should be kept frozen until buffer is added so transfer the sample with pre-cooled spatula and keep the tube in liquid nitrogen all time. Add 350 $\mu$ l preheated extraction buffer (here for 100mg tissue, buffer volume can be as much as 500 $\mu$ l

for bigger samples) to tube, vortex and heat tube to 80°C for ~1 min. Keep then on ice. Vortex sample, grind additionally with electric mortar.

### 3. Digestion

[00125] Add Proteinase K (0.15mg/100mg tissue), vortex and keep at 37°C for one hour.

#### First Purification

[00126] Add 27µl 2M KCl. Chill on ice for 10 min. Centrifuge at 12.000 rpm for 10 minutes at room temperature. Transfer supernatant to fresh, RNAase-free tube and do one phenol extraction, followed by a chloroform:isoamylalcohol extraction. Add 1 vol. isopropanol to supernatant and chill on ice for 10 min. Pellet RNA by centrifugation (7000 rpm for 10 min at RT). Resolve pellet in 1ml 4M LiCl by 10 to 15min vortexing. Pellet RNA by 5min centrifugation.

#### Second Purification

[00127] Resuspend pellet in 500µl Resuspension buffer. Add 500µl phenol and vortex. Add 250µl chloroform:isoamylalcohol and vortex. Spin for 5 min. and transfer supernatant to fresh tube. Repeat chloform:isoamylalcohol extraction until interface is clear. Transfer supernatant to fresh tube and add 1/10 vol 3M NaOAc, pH 5 and 600µl isopropanol. Keep at -20 for 20 min or longer. Pellet RNA by 10 min centrifugation. Wash pellet once with 70% ethanol. Remove all remaining alcohol before resolving pellet with 15 to 20µl DEPC-water. Determine quantity and quality by measuring the absorbance of a 1:200 dilution at 260 and 280nm. 40µg RNA/ml = 1OD<sub>260</sub>

[00128] RNA from wild-type Arabidopsis is isolated as described (Hosein, 2001, *Plant Mol. Biol. Rep.*, **19**, 65a-65e; Ruuska, S.A., Girke, T., Benning, C., & Ohlrogge, J.B., 2002, *Plant Cell*, **14**, 1191-1206).

[00129] The mRNA is prepared from total RNA, using the Amersham Pharmacia Biotech mRNA purification kit, which utilizes oligo(dT)-cellulose columns.

[00130] Isolation of Poly-(A)+ RNA was isolated using Dyna BeadsR (Dyna, Oslo, Norway) following the instructions of the manufacturer's protocol. After determination of the concentration of the RNA or of the poly(A)+ RNA, the RNA was precipitated by addition of 1/10 volumes of 3 M sodium acetate pH 4.6 and 2 volumes of ethanol and stored at -70°C.

***Physcomitrella patens, Brassica napus, Glycine max, Zea mays and Oryza sativa***

[00131] *Physcomitrella patens* was either modified in liquid culture using Knop medium according to Reski & Abel (1985, Planta 165: 354-358) or cultured on Knop solid medium using 1% oxoid agar (Unipath, Basingstoke, England). The protonemas used for RNA and DNA isolation were cultured in aerated liquid cultures. The protonemas were comminuted every 9 days and transferred to fresh culture medium.

[00132] *Brassica napus* and *Glycine max* seeds were separated from pods to create homogeneous materials for seed and seed pod cDNA libraries. Tissues were ground into fine powder under liquid N<sub>2</sub> using a mortar and pestle and transferred to a 50 ml tube. Tissue samples were stored at -80 °C until extractions could be performed. In the case of *Oryza sativa*, 5K - 10K embryos and endosperm were isolated through dissection. Tissues were placed in small tubes or petri dishes on ice during dissection. Containers were placed on dry ice, then stored at -80°C.

[00133] In the case of *Zea mays*, tissues were ground into fine powder under liquid N<sub>2</sub> using a mortar and pestle and transferred to a 50 ml tube. Tissue samples were stored at -80 °C until extractions could be performed.

[00134] Total RNA was extracted from tissues using RNeasy Maxi kit (Qiagen) according to manufacture's protocol and mRNA was processed from total RNA using Oligotex mRNA Purification System kit (Qiagen), also according to manufacture's protocol. mRNA was sent to Hyseq Pharmaceuticals Incorporated (Sunnyville, CA) for further processing of mRNA from each tissue type into cDNA libraries and for use in their proprietary processes, in which similar inserts in plasmids are clustered based on hybridization patterns.

**Example 4: cDNA Library Construction**

[00135] For cDNA library construction, first strand synthesis was achieved using Murine Leukemia Virus reverse transcriptase (Roche, Mannheim, Germany) and oligo-d(T)-primers, second strand synthesis by incubation with DNA polymerase I, Klenow enzyme and RNaseH digestion at 12°C (2 h), 16°C (1 h) and 22°C (1 h). The reaction was stopped by incubation at 65°C (10 min) and subsequently transferred to ice. Double stranded DNA molecules were blunted by T4-DNA-polymerase (Roche, Mannheim) at 37°C (30 min). Nucleotides were removed by phenol/chloroform extraction and Sephadex G50 spin columns. EcoRI adapters (Pharmacia, Freiburg, Germany) were ligated to the cDNA ends by T4-DNA-ligase (Roche, 12°C, overnight) and phosphorylated by incubation with polynucleotide kinase (Roche, 37°C, 30

min). This mixture was subjected to separation on a low melting agarose gel. DNA molecules larger than 300 base pairs were eluted from the gel, phenol extracted, concentrated on Elutip-D-columns (Schleicher and Schuell, Dassel, Germany) and were ligated to vector arms and packed into lambda ZAPII phages or lambda ZAP-Express phages using the Gigapack Gold Kit (Stratagene, Amsterdam, Netherlands) using material and following the instructions of the manufacturer.

[00136] For *Physcomitrella patens* cDNA library construction first strand synthesis was achieved using Murine Leukemia Virus reverse transcriptase (Roche, Mannheim, Germany) and oligo-d(T)-primers, second strand synthesis by incubation with DNA polymerase I, Klenow enzyme and RNaseH digestion at 12°C (2 h), 16°C (1 h) and 22°C (1 h). The reaction was stopped by incubation at 65°C (10 min) and subsequently transferred to ice. Double stranded DNA molecules were blunted by T4-DNA-polymerase (Roche, Mannheim) at 37°C (30 min). Nucleotides were removed by phenol/chloroform extraction and Sephadex G50 spin columns. EcoRI adapters (Pharmacia, Freiburg, Germany) were ligated to the cDNA ends by T4-DNA-ligase (Roche, 12°C, overnight) and phosphorylated by incubation with polynucleotide kinase (Roche, 37°C, 30 min). This mixture was subjected to separation on a low melting agarose gel. DNA molecules larger than 300 basepairs were eluted from the gel, phenol extracted, concentrated on Elutip-D-columns (Schleicher and Schuell, Dassel, Germany) and were ligated to vector arms and packed into lambda ZAPII phages or lambda ZAP-Express phages using the Gigapack Gold Kit (Stratagene, Amsterdam, Netherlands) using material and following the instructions of the manufacturer.

[00137] *Brassica napus*, *Glycine max*, *Zea mays* and *Oryza sativa* cDNA libraries were generated at Hyseq Pharmaceuticals Incorporated (Sunnyville, CA). No amplification steps were used in the library production to retain expression information. Hyseq's genomic approach involves grouping the genes into clusters and then sequencing representative members from each cluster. cDNA libraries were generated from oligo dT column purified mRNA. Colonies from transformation of the cDNA library into *E.coli* were randomly picked and the cDNA insert were amplified by PCR and spotted on nylon membranes. A set of <sup>33</sup>P radiolabeled oligonucleotides were hybridized to the clones and the resulting hybridization pattern determined, to which cluster a particular clone belonged. cDNA clones and their DNA sequences were obtained for use in overexpression in transgenic plants and in other molecular biology processes described herein.

**Example 5: Identification of LMP Genes of Interest that Are *Physcomitrella patens* casein kinase I-like**

***Brassica napus*, *Glycine max*, *Zea mays* and *Oryza sativa***

[00138] This example illustrates how cDNA clones encoding *Physcomitrella patens* casein kinase I-like polypeptides of *Brassica napus*, *Glycine max*, *Zea mays*, and *Oryza sativa* were identified and isolated.

[00139] In order to identify *Physcomitrella patens* casein kinase I-like genes in propriety databases, a similarity analysis using BLAST software (Basic Local Alignment Search Tool, Altschul et al., 1990, J. Mol. Biol. 215:403-410) was carry out. The amino acid sequence of the *Physcomitrella patens* casein kinase I polypeptide was used as a query to search and align DNA databases from *Physcomitrella patens*, *Brassica napus*, *Glycine max*, *Zea mays*, and *Oryza sativa* that were translated in all six reading frames, using the TBLASTN algorithm. Such similarity analysis of the BPS in-house databases resulted in the identification of numerous ESTs and cDNA contigs.

[00140] RNA expression profile data obtained from the Hyseq clustering process were used to determine organ-specificity. Clones showing a greater expression in seed libraries compared to the other tissue libraries were selected as LMP candidate genes. The *Physcomitrella patens*, *Brassica napus*, *Glycine max*, *Zea mays* and *Oryza sativa* clones were selected for overexpression in *Arabidopsis thaliana* cv. Columbia or other crop plants based on their expression profile.

**Example 6: Cloning of full-length cDNAs and orthologs of identified LMP genes**

***Physcomitrella patens*, *Brassica napus*, *Glycine max*, *Zea mays* and *Oryza sativa***

[00141] Full-length sequences of the *Physcomitrella patens*, *Brassica napus*, *Glycine max*, *Zea mays* and *Oryza sativa* partial cDNAs that were identified either in *Physcomitrella patens* EstMax (see also WO 02/074977 A2 for details) or in Hyseq databases are isolated by RACE PCR using the SMART RACE cDNA amplification kit from Clontech allowing both 5'- and 3' rapid amplification of cDNA ends (RACE). The isolation of cDNAs and the RACE PCR protocol used are based on the manufacturer's conditions. The RACE product fragments are extracted from agarose gels with a QIAquick Gel Extraction Kit (Qiagen) and ligated into the TOPO pCR 2.1 vector (Invitrogen) following manufacturer's instructions. Recombinant vectors are transformed into TOP10 cells (Invitrogen) using standard conditions (Sambrook et al. 1989). Transformed cells are grown overnight at 37°C on LB agar containing 50 µg/ml kanamycin and spread with 40µl of a 40 mg/ml stock solution of X-gal in dimethylformamide for blue-white

selection. Single white colonies are selected and used to inoculate 3 ml of liquid LB containing 50 µg/ml kanamycin and grown overnight at 37°C. Plasmid DNA is extracted using the QIAprep Spin Miniprep Kit (Qiagen) following manufacturer's instructions. Subsequent analyses of clones and restriction mapping are performed according to standard molecular biology techniques (Sambrook et al. 1989).

**[00142]** Clones of *Brassica napus*, *Glycine max*, *Zea mays*, or *Oryza sativa* genes obtained from Hyseq were sequenced at using a ABI 377 slab gel sequencer and BigDye Terminator Ready Reaction kits (PE Biosystems, Foster City, CA). Gene specific primers were designed using these sequences and genes were amplified from the plasmid supplied from Hyseq using touch-down PCR. In some cases, primers were designed to add an "AACA" Kozak-like sequence just upstream of the gene start codon and two bases downstream were, in some cases, changed to GC to facilitate increased gene expression levels (Chandrashekhar et al., 1997, Plant Molecular Biology 35:993-1001). PCR reaction cycles were: 94 °C, 5 min; 9 cycles of 94 °C, 1 min, 65 °C, 1 min, 72 °C, 4 min, and in which the anneal temperature was lowered by 1 °C each cycle; 20 cycles of 94 °C, 1 min, 55 °C, 1 min, 72 °C, 4 min; and the PCR cycle was ended with 72 °C, 10 min. Amplified PCR products were gel purified from 1% agarose gels using GenElute -EtBr spin columns (Sigma) and after standard enzymatic digestion, were ligated into the plant binary vector pBPS-GB1 for transformation into *Arabidopsis thaliana* or other crops. The binary vector was amplified by overnight growth in *E. coli* DH5 in LB media and appropriate antibiotic and plasmid was prepared for downstream steps using Qiagen MiniPrep DNA preparation kit. The insert was verified throughout the various cloning steps by determining its size through restriction digest and inserts were sequenced in parallel to plant transformations to ensure the expected gene was used in *Arabidopsis thaliana* or other crop transformation.

**[00143]** Gene sequences can be used to identify homologous or heterologous genes (orthologs, the same LMP gene from another plant) from cDNA or genomic libraries. This can be done by designing PCR primers to conserved sequences identified by multiple sequence alignments. Orthologs are often identified by designing degenerate primers to full-length or partial sequences of genes of interest.

**[00144]** Gene sequences can be used to identify homologues or orthologs from cDNA or genomic libraries. Homologous genes (e. g. full-length cDNA clones) can be isolated via nucleic acid hybridization using for example cDNA libraries: Depending on the abundance of the gene of interest, 100,000 up to 1,000,000 recombinant bacteriophages are plated and transferred to nylon membranes. After denaturation with alkali, DNA is immobilized on the membrane by,

e.g., UV cross linking. Hybridization is carried out at high stringency conditions. Aqueous solution hybridization and washing is performed at an ionic strength of 1 M NaCl and a temperature of 68°C. Hybridization probes are generated by, e.g., radioactive (<sup>32</sup>P) nick transcription labeling (High Prime, Roche, Mannheim, Germany). Signals are detected by autoradiography.

[00145] Partially homologous or heterologous genes that are related but not identical can be identified in a procedure analogous to the above-described procedure using low stringency hybridization and washing conditions. For aqueous hybridization, the ionic strength is normally kept at 1 M NaCl while the temperature is progressively lowered from 68 to 42°C.

[00146] Isolation of gene sequences with homologies (or sequence identity/similarity) only in a distinct domain of (for example 10-20 amino acids) can be carried out by using synthetic radio labeled oligonucleotide probes. Radio labeled oligonucleotides are prepared by phosphorylation of the 5-prime end of two complementary oligonucleotides with T4 polynucleotide kinase. The complementary oligonucleotides are annealed and ligated to form concatemers. The double stranded concatemers are then radiolabeled by for example nick transcription. Hybridization is normally performed at low stringency conditions using high oligonucleotide concentrations.

*Oligonucleotide hybridization solution:*

6 x SSC

0.01 M sodium phosphate

1 mM EDTA (pH 8)

0.5 % SDS

100 µg/ml denaturated salmon sperm DNA

0.1 % nonfat dried milk

[00147] During hybridization, temperature is lowered stepwise to 5-10°C below the estimated oligonucleotide T<sub>m</sub> or down to room temperature followed by washing steps and autoradiography. Washing is performed with low stringency such as 3 washing steps using 4x SSC. Further details are described by Sambrook et al. (1989, "Molecular Cloning: A Laboratory Manual," Cold Spring Harbor Laboratory Press) or Ausubel et al. (1994, "Current Protocols in Molecular Biology," John Wiley & Sons).

**Example 7: Identification of Genes of Interest by Screening Expression Libraries with Antibodies**

[00148] c-DNA clones can be used to produce recombinant protein for example in *E. coli* (e. g. Qiagen QIAexpress pQE system). Recombinant proteins are then normally affinity

purified via Ni-NTA affinity chromatography (Qiagen). Recombinant proteins can be used to produce specific antibodies for example by using standard techniques for rabbit immunization. Antibodies are affinity purified using a Ni-NTA column saturated with the recombinant antigen as described by Gu et al. (1994, *BioTechniques* 17:257-262). The antibody can then be used to screen expression cDNA libraries to identify homologous or heterologous genes via an immunological screening (Sambrook et al., 1989, *Molecular Cloning: A Laboratory Manual*," Cold Spring Harbor Laboratory Press or Ausubel et al., 1994, "Current Protocols in Molecular Biology," John Wiley & Sons).

#### **Example 8: Northern-Hybridization**

[00149] For RNA hybridization, 20 µg of total RNA or 1 µg of poly-(A)<sup>+</sup> RNA is separated by gel electrophoresis in 1.25% agarose gels using formaldehyde as described in Amasino (1986, *Anal. Biochem.* 152:304), transferred by capillary attraction using 10 x SSC to positively charged nylon membranes (Hybond N<sup>+</sup>, Amersham, Braunschweig), immobilized by UV light and pre-hybridized for 3 hours at 68°C using hybridization buffer (10% dextran sulfate w/v, 1 M NaCl, 1% SDS, 100 µg/ml of herring sperm DNA). The labeling of the DNA probe with the Highprime DNA labeling kit (Roche, Mannheim, Germany) is carried out during the pre-hybridization using alpha-32P dCTP (Amersham, Braunschweig, Germany). Hybridization is carried out after addition of the labeled DNA probe in the same buffer at 68°C overnight. The washing steps are carried out twice for 15 min using 2 x SSC and twice for 30 min using 1 x SSC, 1% SDS at 68°C. The exposure of the sealed filters is carried out at -70°C for a period of 1 day to 14 days.

#### **Example 9: DNA Sequencing and Computational Functional Analysis**

[00150] cDNA libraries were used for DNA sequencing according to standard methods, in particular by the chain termination method using the ABI PRISM Big Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer, Weiterstadt, Germany). Random sequencing was carried out subsequent to preparative plasmid recovery from cDNA libraries via *in vivo* mass excision, retransformation, and subsequent plating of DH10B on agar plates (material and protocol details from Stratagene, Amsterdam, Netherlands). Plasmid DNA was prepared from overnight grown *E. coli* cultures grown in Luria-Broth medium containing ampicillin (see Sambrook et al. (1989, Cold Spring Harbor Laboratory Press: ISBN 0-87969-309-6) on a

Qiagen DNA preparation robot (Qiagen, Hilden) according to the manufacturer's protocols). Sequencing primers with the following nucleotide sequences were used:

5'-CAGGAAACAGCTATGACC-3'  
5'-CTAAAGGGAACAAAAGCTG-3'  
5'-TGTAACGACGGCCAGT-3'

[00151] Sequences were processed and annotated using the software package EST-MAX commercially provided by Bio-Max (Munich, Germany). The program incorporates bioinformatics methods important for functional and structural characterization of protein sequences. For reference see <http://pedant.mips.biochem.mpg.de>.

[00152] The most important algorithms incorporated in EST-MAX are: FASTA: Very sensitive protein sequence database searches with estimates of statistical significance (Pearson W.R. 1990, Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol.* 183:63-98). BLAST: Very sensitive protein sequence database searches with estimates of statistical significance (Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; and Lipman, D.J., "Basic local alignment search tool," *J. Mol. Biol.* 215:403-410). PREDATOR: High-accuracy secondary structure prediction from single and multiple sequences. (Frishman & Argos, 1997, 75% accuracy in protein secondary structure prediction. *Proteins* 27:329-335). CLUSTALW: Multiple sequence alignment (Thompson, J.D., Higgins, D.G., and Gibson, T.J., 1994, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice, *Nucleic Acids Res.* 22:4673-4680). TMAP: Transmembrane region prediction from multiply aligned sequences (Persson B. & Argos P., 1994, Prediction of transmembrane segments in proteins utilizing multiple sequence alignments, *J. Mol. Biol.* 237:182-192). ALOM2: Transmembrane region prediction from single sequences (Klein P., Kanehisa M., and DeLisi C., 1984, "Prediction of protein function from sequence properties: A discriminant analysis of a database," *Biochim. Biophys. Acta* 787:221-226. Version 2 by Dr. K. Nakai). PROSEARCH: Detection of PROSITE protein sequence patterns. Kolakowski, L.F. Jr., Leunissen, J.A.M., and Smith, J.E. 1992, ProSearch, "fast searching of protein sequences with regular expression patterns related to protein structure and function," *Biotechniques* 13:919-921). BLIMPS: Similarity searches against a database of ungapped blocks (Wallace & Henikoff, 1992, PATMAT: "A searching and extraction program for sequence, pattern and block queries and databases," *CABIOS* 8:249-254. Written by Bill Alford).

**Example 10: Plasmids for Plant Transformation**

[00153] For plant transformation binary vectors such as pBinAR can be used (Höfgen & Willmitzer, 1990, Plant Sci. 66:221-230). Construction of the binary vectors can be performed by ligation of the cDNA in sense or antisense orientation into the T-DNA. 5-prime to the cDNA a plant promoter activates transcription of the cDNA. A polyadenylation sequence is located 3'-prime to the cDNA. Tissue-specific expression can be achieved by using a tissue specific promoter. For example, seed-specific expression can be achieved by cloning the napin or LeB4 or USP promoter 5-prime to the cDNA. Also any other seed specific promoter element can be used. For constitutive expression within the whole plant the CaMV 35S promoter can be used. The expressed protein can be targeted to a cellular compartment using a signal peptide, for example for plastids, mitochondria, or endoplasmic reticulum (Kermode 1996, Crit. Rev. Plant Sci. 15:285-423). The signal peptide is cloned 5-prime in frame to the cDNA to achieve subcellular localization of the fusion protein.

[00154] Further examples for plant binary vectors are the pBPS-GB1 or pSUN2-GW vectors into which the LMP gene candidates are cloned. These binary vectors contain an antibiotic resistance gene driven under the control of the AtAct2-I promoter or the Nos-promotor, respectively, and a USP seed-specific promoter in front of the candidate gene with the NOSpA terminator or the OCS terminator. Partial or full-length LMP cDNA are cloned into the multiple cloning site of the plant binary vector in sense or antisense orientation behind the USP seed-specific promoter. The recombinant vector containing the gene of interest is transformed into Top10 cells (Invitrogen) using standard conditions. Transformed cells are selected for on LB agar containing 50 µg/ml kanamycin grown overnight at 37°C. Plasmid DNA is extracted using the QIAprep Spin Miniprep Kit (Qiagen) following manufacturer's instructions. Analysis of subsequent clones and restriction mapping is performed according to standard molecular biology techniques (Sambrook et al., 1989, "Molecular Cloning, A Laboratory Manual," 2<sup>nd</sup> Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

**Example 11: *Agrobacterium* Mediated Plant Transformation**

[00155] *Agrobacterium* mediated plant transformation with the LMP nucleic acids described herein can be performed using standard transformation and regeneration techniques (Gelvin, Stanton B. & Schilperoort R.A, Plant Molecular Biology Manual, 2nd ed. Kluwer Academic Publ., Dordrecht, 1995 in Sect., Ringbuc Zentrale Signatur:BT11-P; Glick, Bernard R. and Thompson, John E. Methods in Plant Molecular Biology and Biotechnology, S. 360, CRC

Press, Boca Raton 1993). For example, *Agrobacterium* mediated transformation can be performed using the GV3 (pMP90) (Koncz & Schell, 1986, Mol. Gen. Genet. 204:383-396) or LBA4404 (Clontech) *Agrobacterium tumefaciens* strain.

[00156] *Arabidopsis thaliana* can be grown and transformed according to standard conditions (Bechtold 1993, Acad. Sci. Paris. 316:1194-1199; Bent et al. 1994, Science 265:1856-1860). Additionally, rapeseed can be transformed with the LMR nucleic acids of the present invention via cotyledon or hypocotyl transformation (Moloney et al. 1989, Plant Cell Report 8:238-242; De Block et al. 1989, Plant Physiol. 91:694-701). Use of antibiotic for *Agrobacterium* and plant selection depends on the binary vector and the *Agrobacterium* strain used for transformation. Rapeseed selection is normally performed using a selectable plant marker. Additionally, *Agrobacterium* mediated gene transfer to flax can be performed using, for example, a technique described by Mlynarova et al. (1994, Plant Cell Report 13:282-285).

[00157] Transformation of soybean can be performed using for example a technique described in EP 0424 047, U.S. Patent No. 5,322,783 (Pioneer Hi-Bred International) or in EP 0397 687, U.S. Patent No. 5,376,543 or U.S. Patent No. 5,169,770 (University Toledo), or by any of a number of other transformation procedures known in the art. Soybean seeds are surface sterilized with 70% ethanol for 4 minutes at room temperature with continuous shaking, followed by 20% (v/v) Clorox supplemented with 0.05% (v/v) Tween for 20 minutes with continuous shaking. Then the seeds are rinsed 4 times with distilled water and placed on moistened sterile filter paper in a Petri dish at room temperature for 6 to 39 hours. The seed coats are peeled off, and cotyledons are detached from the embryo axis. The embryo axis is examined to make sure that the meristematic region is not damaged. The excised embryo axes are collected in a half-open sterile Petri dish and air-dried to a moisture content less than 20% (fresh weight) in a sealed Petri dish until further use.

[00158] The method of plant transformation is also applicable to *Brassica napus* and other crops. In particular, seeds of canola are surface sterilized with 70% ethanol for 4 minutes at room temperature with continuous shaking, followed by 20% (v/v) Clorox supplemented with 0.05 % (v/v) Tween for 20 minutes, at room temperature with continuous shaking. Then, the seeds are rinsed 4 times with distilled water and placed on moistened sterile filter paper in a Petri dish at room temperature for 18 hours. The seed coats are removed and the seeds are air dried overnight in a half-open sterile Petri dish. During this period, the seeds lose approximately 85% of their water content. The seeds are then stored at room temperature in a sealed Petri dish until further use.

[00159] *Agrobacterium tumefaciens* culture is prepared from a single colony in LB solid medium plus appropriate antibiotics (e.g. 100 mg/l streptomycin, 50 mg/l kanamycin) followed by growth of the single colony in liquid LB medium to an optical density at 600 nm of 0.8. Then, the bacteria culture is pelleted at 7000 rpm for 7 minutes at room temperature, and re-suspended in MS (Murashige & Skoog 1962, *Physiol. Plant.* 15:473-497) medium supplemented with 100 mM acetosyringone. Bacteria cultures are incubated in this pre-induction medium for 2 hours at room temperature before use. The axis of soybean zygotic seed embryos at approximately 44% moisture content are imbibed for 2 h at room temperature with the pre-induced *Agrobacterium* suspension culture. (The imbibition of dry embryos with a culture of *Agrobacterium* is also applicable to maize embryo axes.)

[00160] The embryos are removed from the imbibition culture and are transferred to Petri dishes containing solid MS medium supplemented with 2% sucrose and incubated for 2 days, in the dark at room temperature. Alternatively, the embryos are placed on top of moistened (liquid MS medium) sterile filter paper in a Petri dish and incubated under the same conditions described above. After this period, the embryos are transferred to either solid or liquid MS medium supplemented with 500 mg/l carbenicillin or 300 mg/l cefotaxime to kill the agrobacteria. The liquid medium is used to moisten the sterile filter paper. The embryos are incubated during 4 weeks at 25°C, under 440  $\mu\text{mol m}^{-2}\text{s}^{-1}$  and 12 hours photoperiod. Once the seedlings have produced roots, they are transferred to sterile metromix soil. The medium of the *in vitro* plants is washed off before transferring the plants to soil. The plants are kept under a plastic cover for 1 week to favor the acclimatization process. Then the plants are transferred to a growth room where they are incubated at 25°C, under 440  $\mu\text{mol m}^{-2}\text{s}^{-1}$  light intensity and 12 h photoperiod for about 80 days.

[00161] Samples of the primary transgenic plants ( $T_0$ ) are analyzed by PCR to confirm the presence of T-DNA. These results are confirmed by Southern hybridization wherein DNA is electrophoresed on a 1% agarose gel and transferred to a positively charged nylon membrane (Roche Diagnostics). The PCR DIG Probe Synthesis Kit (Roche Diagnostics) is used to prepare a digoxigenin-labeled probe by PCR as recommended by the manufacturer.

#### Example 12: *In vivo* Mutagenesis

[00162] *In vivo* mutagenesis of microorganisms can be performed by incorporation and passage of the plasmid (or other vector) DNA through *E. coli* or other microorganisms (e.g. *Bacillus* spp. or yeasts such as *Saccharomyces cerevisiae*), which are impaired in their

capabilities to maintain the integrity of their genetic information. Typical mutator strains have mutations in the genes for the DNA repair system (e.g., mutHLS, mutD, mutT, etc.; for reference, see Rupp W.D., 1996, DNA repair mechanisms, in: *Escherichia coli* and *Salmonella*, p. 2277-2294, ASM: Washington.) Such strains are well known to those skilled in the art. The use of such strains is illustrated, for example, in Greener and Callahan 1994, Strategies 7:32-34. Transfer of mutated DNA molecules into plants is preferably done after selection and testing in microorganisms. Transgenic plants are generated according to various examples within the exemplification of this document.

**Example 13: Assessment of the mRNA Expression and Activity of a Recombinant Gene Product in the Transformed Organism**

[00163] The activity of a recombinant gene product in the transformed host organism can be measured on the transcriptional or/and on the translational level. A useful method to ascertain the level of transcription of the gene (an indicator of the amount of mRNA available for translation to the gene product) is to perform a Northern blot (for reference see, for example, Ausubel et al. 1988, Current Protocols in Molecular Biology, Wiley: New York), in which a primer designed to bind to the gene of interest is labeled with a detectable tag (usually radioactive or chemiluminescent), such that when the total RNA of a culture of the organism is extracted, run on gel, transferred to a stable matrix and incubated with this probe, the binding and quantity of binding of the probe indicates the presence and also the quantity of mRNA for this gene. This information at least partially demonstrates the degree of transcription of the transformed gene. Total cellular RNA can be prepared from plant cells, tissues, or organs by several methods, all well-known in the art, such as that described in Bormann et al. (1992, Mol. Microbiol. 6:317-326).

[00164] To assess the presence or relative quantity of protein translated from this mRNA, standard techniques, such as a Western blot, may be employed (see, for example, Ausubel et al. 1988, Current Protocols in Molecular Biology, Wiley: New York). In this process, total cellular proteins are extracted, separated by gel electrophoresis, transferred to a matrix such as nitrocellulose, and incubated with a probe, such as an antibody, which specifically binds to the desired protein. This probe is generally tagged with a chemiluminescent or colorimetric label, which may be readily detected. The presence and quantity of label observed indicates the presence and quantity of the desired mutant protein present in the cell.

[00165] The activity of LMPs that bind to DNA can be measured by several well-established methods, such as DNA band-shift assays (also called gel retardation assays). The

effect of such LMP on the expression of other molecules can be measured using reporter gene assays (such as that described in Kolmar H. et al. 1995, EMBO J. 14:3895-3904 and references cited therein). Reporter gene test systems are well known and established for applications in both prokaryotic and eukaryotic cells, using enzymes such as beta-galactosidase, green fluorescent protein, and several others.

[00166] The determination of activity of lipid metabolism membrane-transport proteins can be performed according to techniques such as those described in Gennis R.B. (1989 Pores, Channels and Transporters, in Biomembranes, Molecular Structure and Function, Springer: Heidelberg, pp. 85-137, 199-234 and 270-322).

**Example 14: *In vitro* Analysis of the Function of *Physcomitrella patens*, *Brassica napus*, *Glycine max*, *Zea mays* and *Oryza sativa* casein kinase-like Genes in Transgenic Plants**

[00167] The determination of activities and kinetic parameters of enzymes is well established in the art. Experiments to determine the activity of any given altered enzyme must be tailored to the specific activity of the wild-type enzyme, which is well within the ability of one skilled in the art. Overviews about enzymes in general, as well as specific details concerning structure, kinetics, principles, methods, applications and examples for the determination of many enzyme activities may be found, for example, in the following references: Dixon, M. & Webb, E.C. 1979, Enzymes. Longmans: London; Fersht, (1985) Enzyme Structure and Mechanism. Freeman: New York; Walsh (1979) Enzymatic Reaction Mechanisms. Freeman: San Francisco; Price, N.C., Stevens, L. (1982) Fundamentals of Enzymology. Oxford Univ. Press: Oxford; Boyer, P.D., ed. (1983) The Enzymes, 3rd ed. Academic Press: New York; Bisswanger, H., (1994) Enzymkinetik, 2nd ed. VCH: Weinheim (ISBN 3527300325); Bergmeyer, H.U., Bergmeyer, J., Graßl, M., eds. (1983-1986) Methods of Enzymatic Analysis, 3rd ed., vol. I-XII, Verlag Chemie: Weinheim; and Ullmann's Encyclopedia of Industrial Chemistry (1987) vol. A9, Enzymes. VCH: Weinheim, p. 352-363.

**Example 15: Analysis of the Impact of Recombinant Proteins on the Production of a Desired Seed Storage Compound**

[00168] The effect of the genetic modification in plants on a desired seed storage compound (such as a sugar, lipid, or fatty acid) can be assessed by growing the modified plant under suitable conditions and analyzing the seeds or any other plant organ for increased production of the desired product (i.e., a lipid or a fatty acid). Such analysis techniques are well known to one skilled in the art, and include spectroscopy, thin layer chromatography, staining

methods of various kinds, enzymatic and microbiological methods, and analytical chromatography such as high performance liquid chromatography (see, for example, Ullman, 1985, Encyclopedia of Industrial Chemistry, vol. A2, pp. 89-90 and 443-613, VCH: Weinheim; Fallon, A. et al., 1987, Applications of HPLC in Biochemistry in: Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17; Rehm et al., 1993 Product recovery and purification, Biotechnology, vol. 3, Chapter III, pp. 469-714, VCH: Weinheim; Belter, P.A. et al., 1988 Bioseparations: downstream processing for biotechnology, John Wiley & Sons; Kennedy J.F. & Cabral J.M.S. 1992, Recovery processes for biological materials, John Wiley and Sons; Shaeiwitz J.A. & Henry J.D. 1988, Biochemical separations in: Ulmann's Encyclopedia of Industrial Chemistry, Separation and purification techniques in biotechnology, vol. B3, Chapter 11, pp. 1-27, VCH: Weinheim; and Dechow F.J. 1989).

**[00169]** Besides the above-mentioned methods, plant lipids are extracted from plant material as described by Cahoon et al. (1999, Proc. Natl. Acad. Sci. USA 96, 22:12935-12940) and Browse et al. (1986, Anal. Biochemistry 442:141-145). Qualitative and quantitative lipid or fatty acid analysis is described in Christie, William W., Advances in Lipid Methodology. Ayr/Scotland :Oily Press. - (Oily Press Lipid Library; Christie, William W., Gas Chromatography and Lipids. A Practical Guide - Ayr, Scotland :Oily Press, 1989 Repr. 1992. - IX,307 S. - (Oily Press Lipid Library; and "Progress in Lipid Research, Oxford :Pergamon Press, 1 (1952) – 16 (1977) Progress in the Chemistry of Fats and Other Lipids CODEN.

**[00170]** Unequivocal proof of the presence of fatty acid products can be obtained by the analysis of transgenic plants following standard analytical procedures: GC, GC-MS or TLC as variously described by Christie and references therein (1997 in: Advances on Lipid Methodology 4th ed.: Christie, Oily Press, Dundee, pp. 119-169; 1998). Detailed methods are described for leaves by Lemieux et al. (1990, Theor. Appl. Genet. 80:234-240) and for seeds by Focks & Benning (1998, Plant Physiol. 118:91-101).

**[00171]** Positional analysis of the fatty acid composition at the sn-1, sn-2 or sn-3 positions of the glycerol backbone is determined by lipase digestion (see, e.g., Siebertz & Heinz 1977, Z. Naturforsch. 32c:193-205, and Christie 1987, Lipid Analysis 2<sup>nd</sup> Edition, Pergamon Press, Exeter, ISBN 0-08-023791-6).

**[00172]** A typical way to gather information regarding the influence of increased or decreased protein activities on lipid and sugar biosynthetic pathways is for example via analyzing the carbon fluxes by labeling studies with leaves or seeds using <sup>14</sup>C-acetate or <sup>14</sup>C-pyruvate (see, e.g. Focks & Benning 1998, Plant Physiol. 118:91-101; Eccleston & Ohlrogge

1998, *Plant Cell* 10:613-621). The distribution of carbon-14 into lipids and aqueous soluble components can be determined by liquid scintillation counting after the respective separation (for example on TLC plates) including standards like  $^{14}\text{C}$ -sucrose and  $^{14}\text{C}$ -malate (Eccleston & Ohlrogge 1998, *Plant Cell* 10:613-621).

[00173] Material to be analyzed can be disintegrated via sonification, glass milling, liquid nitrogen and grinding or via other applicable methods. The material has to be centrifuged after disintegration. The sediment is re-suspended in distilled water, heated for 10 minutes at 100°C, cooled on ice and centrifuged again followed by extraction in 0.5 M sulfuric acid in methanol containing 2% dimethoxypropane for 1 hour at 90°C leading to hydrolyzed oil and lipid compounds resulting in transmethylated lipids. These fatty acid methyl esters are extracted in petrolether and finally subjected to GC analysis using a capillary column (Chrompack, WCOT Fused Silica, CP-Wax-52 CB, 25 m, 0.32 mm) at a temperature gradient between 170°C and 240°C for 20 minutes and 5 min. at 240°C. The identity of resulting fatty acid methylesters is defined by the use of standards available from commercial sources (i.e., Sigma).

[00174] In case of fatty acids where standards are not available, molecule identity is shown via derivatization and subsequent GC-MS analysis. For example, the localization of triple bond fatty acids is shown via GC-MS after derivatization via 4,4-Dimethoxy-oxazolin-Derivaten (Christie, Oily Press, Dundee, 1998).

[00175] A common standard method for analyzing sugars, especially starch, is published by Stitt M., Lilley R.Mc.C., Gerhardt R. and Heldt M.W. (1989, "Determination of metabolite levels in specific cells and subcellular compartments of plant leaves" *Methods Enzymol.* 174:518-552; for other methods see also Härtel et al., 1998, *Plant Physiol. Biochem.* 36:407-417 and Focks & Benning 1998, *Plant Physiol.* 118:91-101).

[00176] For the extraction of soluble sugars and starch, 50 seeds are homogenized in 500  $\mu\text{l}$  of 80% (v/v) ethanol in a 1.5-ml polypropylene test tube and incubated at 70°C for 90 min. Following centrifugation at 16,000 g for 5 min, the supernatant is transferred to a new test tube. The pellet is extracted twice with 500  $\mu\text{l}$  of 80% ethanol. The solvent of the combined supernatants is evaporated at room temperature under a vacuum. The residue is dissolved in 50  $\mu\text{l}$  of water, representing the soluble carbohydrate fraction. The pellet left from the ethanol extraction, which contains the insoluble carbohydrates including starch, is homogenized in 200  $\mu\text{l}$  of 0.2 N KOH, and the suspension is incubated at 95°C for 1 h to dissolve the starch. Following the addition of 35  $\mu\text{l}$  of 1 N acetic acid and centrifugation for 5 min at 16,000 g, the supernatant is used for starch quantification.

[00177] To quantify soluble sugars, 10  $\mu$ l of the sugar extract is added to 990  $\mu$ l of reaction buffer containing 100 mM imidazole, pH 6.9, 5 mM MgCl<sub>2</sub>, 2 mM NADP, 1 mM ATP, and 2 units 2 ml<sup>-1</sup> of Glucose-6-P-dehydrogenase. For enzymatic determination of glucose, fructose and sucrose, 4.5 units of hexokinase, 1 unit of phosphoglucosomerase, and 2  $\mu$ l of a saturated fructosidase solution are added in succession. The production of NADPH is photometrically monitored at a wavelength of 340 nm. Similarly, starch is assayed in 30  $\mu$ l of the insoluble carbohydrate fraction with a kit from Boehringer Mannheim.

[00178] An example for analyzing the protein content in leaves and seeds can be found by Bradford M.M. (1976, "A rapid and sensitive method for the quantification of microgram quantities of protein using the principle of protein dye binding" *Anal. Biochem.* 72:248-254). For quantification of total seed protein, 15-20 seeds are homogenized in 250  $\mu$ l of acetone in a 1.5-ml polypropylene test tube. Following centrifugation at 16,000 g, the supernatant is discarded and the vacuum-dried pellet is resuspended in 250  $\mu$ l of extraction buffer containing 50 mM Tris-HCl, pH 8.0, 250 mM NaCl, 1 mM EDTA, and 1% (w/v) SDS. Following incubation for 2 h at 25°C, the homogenate is centrifuged at 16,000 g for 5 min and 200  $\mu$ l of the supernatant will be used for protein measurements. In the assay,  $\gamma$ -globulin is used for calibration. For protein measurements, Lowry DC protein assay (Bio-Rad) or Bradford-assay (Bio-Rad) is used.

[00179] Enzymatic assays of hexokinase and fructokinase are performed spectrophotometrically according to Renz et al. (1993, *Planta* 190:156-165), of phosphogluco-isomerase, ATP-dependent 6-phosphofructokinase, pyrophosphate-dependent 6-phospho-fructokinase, Fructose-1,6-bisphosphate aldolase, triose phosphate isomerase, glyceral-3-P dehydrogenase, phosphoglycerate kinase, phosphoglycerate mutase, enolase and pyruvate kinase are performed according to Burrell et al. (1994, *Planta* 194:95-101) and of UDP-Glucose-pyrophosphorylase according to Zrenner et al. (1995, *Plant J.* 7:97-107).

[00180] Intermediates of the carbohydrate metabolism, like Glucose-1-phosphate, Glucose-6-phosphate, Fructose-6-phosphate, Phosphoenolpyruvate, Pyruvate, and ATP are measured as described in Härtel et al. (1998, *Plant Physiol. Biochem.* 36:407-417) and metabolites are measured as described in Jelitto et al. (1992, *Planta* 188:238-244).

[00181] In addition to the measurement of the final seed storage compound (i.e., lipid, starch or storage protein), it is also possible to analyze other components of the metabolic pathways utilized for the production of a desired seed storage compound, such as intermediates

and side-products, to determine the overall efficiency of production of the compound (Fiehn et al. 2000, Nature Biotech. 18:1447-1161).

[00182] For example, yeast expression vectors comprising the nucleic acids disclosed herein, or fragments thereof, can be constructed and transformed into *Saccharomyces cerevisiae* using standard protocols. The resulting transgenic cells can then be assayed for alterations in sugar, oil, lipid, or fatty acid contents.

[00183] Similarly, plant expression vectors comprising the nucleic acids disclosed herein, or fragments thereof, can be constructed and transformed into an appropriate plant cell such as Arabidopsis, soybean, rapeseed, rice, maize, wheat, *Medicago truncatula*, etc., using standard protocols. The resulting transgenic cells and/or plants derived there from can then be assayed for alterations in sugar, oil, lipid or fatty acid contents.

[00184] Additionally, the sequences disclosed herein, or fragments thereof, can be used to generate knockout mutations in the genomes of various organisms, such as bacteria, mammalian cells, yeast cells, and plant cells (Girke et al. 1998, Plant J. 15:39-48). The resultant knockout cells can then be evaluated for their composition and content in seed storage compounds, and the effect on the phenotype and/or genotype of the mutation. For other methods of gene inactivation include US 6004804 "Non-Chimeric Mutational Vectors" and Puttaraju et al. (1999, "Spliceosome-mediated RNA *trans*-splicing as a tool for gene therapy" Nature Biotech. 17:246-252).

#### **Example 16: Purification of the Desired Product from *Transformed* Organisms**

[00185] An LMP can be recovered from plant material by various methods well known in the art. Organs of plants can be separated mechanically from other tissue or organs prior to isolation of the seed storage compound from the plant organ. Following homogenization of the tissue, cellular debris is removed by centrifugation and the supernatant fraction containing the soluble proteins is retained for further purification of the desired compound. If the product is secreted from cells grown in culture, then the cells are removed from the culture by low-speed centrifugation and the supernate fraction is retained for further purification.

[00186] The supernatant fraction from either purification method is subjected to chromatography with a suitable resin, in which the desired molecule is either retained on a chromatography resin while many of the impurities in the sample are not, or where the impurities are retained by the resin, while the sample is not. Such chromatography steps may be repeated as necessary, using the same or different chromatography resins. One skilled in the art would be

well-versed in the selection of appropriate chromatography resins and in their most efficacious application for a particular molecule to be purified. The purified product may be concentrated by filtration or ultrafiltration, and stored at a temperature at which the stability of the product is maximized.

[00187] There is a wide array of purification methods known to the art and the preceding method of purification is not meant to be limiting. Such purification techniques are described, for example, in Bailey J.E. & Ollis D.F. 1986, *Biochemical Engineering Fundamentals*, McGraw-Hill:New York).

[00188] The identity and purity of the isolated compounds may be assessed by techniques standard in the art. These include high-performance liquid chromatography (HPLC), spectroscopic methods, staining methods, thin layer chromatography, analytical chromatography such as high performance liquid chromatography, NIRS, enzymatic assay, or microbiologically. Such analysis methods are reviewed in: Patek et al. (1994, *Appl. Environ. Microbiol.* 60:133-140), Malakhova et al. (1996, *Biotekhnologiya* 11:27-32) and Schmidt et al. (1998, *Bioprocess Engineer* 19:67-70), *Ulmann's Encyclopedia of Industrial Chemistry* (1996, Vol. A27, VCH: Weinheim, p. 89-90, p. 521-540, p. 540-547, p. 559-566, 575-581 and p. 581-587) and Michal G. (1999, *Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology*, John Wiley and Sons; Fallon, A. et al. 1987, *Applications of HPLC in Biochemistry in: Laboratory Techniques in Biochemistry and Molecular Biology*, vol. 17).

#### Example 17: Screening for Increased Stress Tolerance and Plant Growth

[00189] The transgenic plants are screened for their improved stress tolerance demonstrating that transgene expression confers stress tolerance.

[00190] The transgenic plants are further screened for their growth rate demonstrating that transgene expression confers increased growth rates and/or increased seed yield.

**Table 1.**  
**Plant Lipid Classes**

Neutral Lipids	Triacylglycerol (TAG)
	Diacylglycerol (DAG)
	Monoacylglycerol (MAG)
Polar Lipids	Monogalactosyldiacylglycerol (MGDG)
	Digalactosyldiacylglycerol (DGDG)
	Phosphatidylglycerol (PG)
	Phosphatidylcholine (PC)

	Phosphatidylethanolamine (PE)
	Phosphatidylinositol (PI)
	Phosphatidylserine (PS)
	Sulfoquinovosyldiacylglycerol

**Table 2.**  
**Common Plant Fatty Acids**

16:0	Palmitic acid
16:1	Palmitoleic acid
16:3	Palmitolenic acid
18:0	Stearic acid
18:1	Oleic acid
18:2	Linoleic acid
18:3	Linolenic acid
$\gamma$ -18:3	Gamma-linolenic acid*
20:0	Arachidic acid
20:1	Eicosenoic acid
22:6	Docosahexanoic acid (DHA) *
20:2	Eicosadienoic acid
20:4	Arachidonic acid (AA) *
20:5	Eicosapentaenoic acid (EPA) *
22:1	Erucic acid

\* These fatty acids do not normally occur in plant seed oils, but their production in transgenic plant seed oil is of importance in plant biotechnology.

**Table 3.**  
**A table of the putative functions of the *Physcomitrella patens* casein kinase I-like LMPs (the full length nucleic acid sequences can be found in Appendix A using the sequence codes)**

Species	Sequence ID	Function	ORF length (with stop codon)
<i>Physcomitrella patens</i>	PpCk63	casein kinase I	1422
<i>Brassica napus</i>	BnCk01	casein kinase II alpha-type chain	999
<i>Glycine max</i>	GmCk02	casein kinase II alpha chain	1044
<i>Zea mays</i>		Protein kinase CK2 catalytic	
	ZmCk01	subunit CK2 alpha-3	999
<i>Oryza sativa</i>	OsCk01	casein kinase II alpha subunit	1002

[00191] Those skilled in the art will recognize, or will be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention

described herein. Such equivalents are intended to be encompassed by the claims to the invention disclosed and claimed herein.

**What is claimed is:**

1. A method of producing a transgenic plant having a modified level of a seed storage compound comprising, transforming a plant cell with an expression vector comprising a lipid metabolism protein (LMP) nucleic acid and generating from the plant cell the transgenic plant, wherein the nucleic acid encodes a polypeptide that functions as a modulator of a seed storage compound in the plant, and wherein the LMP nucleic acid comprises a polynucleotide sequence selected from the group consisting of: the sequences shown in Figure 1B, Figure 2B, Figure 3B, Figure 4B, and Figure 5B.
2. A method of producing a transgenic plant having a modified level of a seed storage compound comprising, transforming a plant cell with an expression vector comprising a LMP nucleic acid and generating from the plant cell the transgenic plant, wherein the nucleic acid encodes a polypeptide that functions as a modulator of a seed storage compound in the plant, and wherein the LMP nucleic acid comprises a polynucleotide sequence encoding a polypeptide selected from the group consisting of: Figure 1C, Figure 2C, Figure 3C, Figure 4C, and Figure 5C.
3. A method of producing a transgenic plant having a modified level of a seed storage compound comprising, transforming a plant cell with an expression vector comprising a LMP nucleic acid and generating from the plant cell the transgenic plant, wherein the nucleic acid encodes a polypeptide that functions as a modulator of a seed storage compound in the plant, and wherein the LMP nucleic acid comprises a polynucleotide of least 60 consecutive nucleotides of the LMP nucleic acid of claim 1 or claim 2.
4. A method of producing a transgenic plant having a modified level of a seed storage compound comprising, transforming a plant cell with an expression vector comprising a LMP nucleic acid and generating from the plant cell the transgenic plant, wherein the nucleic acid encodes a polypeptide that functions as a modulator of a seed storage compound in the plant, and wherein the LMP nucleic acid comprises a polynucleotide having at least 70% sequence identity with the LMP nucleic acid of claim 1 or claim 2.
5. A method of producing a transgenic plant having a modified level of a seed storage compound comprising, transforming a plant cell with an expression vector comprising a LMP nucleic acid and generating from the plant cell the transgenic plant, wherein the nucleic acid

encodes a polypeptide that functions as a modulator of a seed storage compound in the plant, and wherein the LMP nucleic acid comprises a polynucleotide having at least 90% sequence identity with the LMP nucleic acid of claim 1 or claim 2.

6. A method of producing a transgenic plant having a modified level of a seed storage compound comprising, transforming a plant cell with an expression vector comprising a LMP nucleic acid and generating from the plant cell the transgenic plant, wherein the nucleic acid encodes a polypeptide that functions as a modulator of a seed storage compound in the plant, and wherein the LMP nucleic acid comprises a first nucleic acid that hybridizes under stringent conditions to a second nucleic acid nucleic acid of claim 1 or claim 2.

7. A method of producing a transgenic plant having a modified level of a seed storage compound comprising, transforming a plant cell with an expression vector comprising a LMP nucleic acid and generating from the plant cell the transgenic plant, wherein the nucleic acid encodes a polypeptide that functions as a modulator of a seed storage compound in the plant, and wherein the LMP nucleic acid comprises a polynucleotide complementary to the LMP nucleic acid of claim 1 or claim 2.

8. A method of modulating a level of a seed storage compound in a plant comprising, modifying the expression of a LMP nucleic acid in the plant, wherein the LMP nucleic acid is selected from the group consisting of the LMP nucleic acids of claims 1, 2, 3, 4, 5, 6 or 7.

9. The method of claim 8, wherein the plant is transgenic.

10. The method of claim 8, wherein the plant is not transgenic.

11. The method of claims 1, 2, 3, 4, 5, 6 or 8, wherein the nucleic acid encodes a polypeptide that contains a protein kinase domain.

12. The method of claim 11, wherein the nucleic acid encodes a polypeptide of Figure 1C, Figure 2C, Figure 3C, Figure 4C, and Figure 5C.

13. A transgenic plant made by the method of claims 1, 2, 3, 4, 5, 6 or 7 wherein expression of the LMP nucleic acid in the plant results in a modified level of a seed storage compound in the plant as compared to a wild type variety of the plant.
14. The transgenic plant of claim 13, wherein the plant is a dicotyledonous plant.
15. The transgenic plant of claim 13, wherein the plant is a monocotyledonous plant.
16. The transgenic plant of claim 13, wherein the plant is an oil producing species.
17. The transgenic plant of claim 13, wherein the plant is selected from the group consisting of rapeseed, canola, linseed, soybean, sunflower, maize, oat, rye, barley, wheat, sugarbeet, tagetes, cotton, oil palm, coconut palm, flax, castor and peanut.
18. The transgenic plant of claim 13, wherein the level of the seed storage compound is increased.
19. The transgenic plant of claim 13, wherein the seed storage compound is selected from the group consisting of a lipid, a fatty acid, a starch and a seed storage protein.
20. A seed produced by the transgenic plant of claim 13, wherein the plant is true breeding for a modified level of the seed storage compound as compared to a wild type variety of the plant.
21. A seed oil produced by the seed of claim 13.
22. An isolated LMP nucleic acid comprising a polynucleotide sequence encoding a polypeptide that functions as a modulator of a seed storage compound in a plant, wherein the polynucleotide is selected from the group consisting of the sequences shown in Figure 1B, Figure 2B, Figure 3B, Figure 4B, and Figure 5B.
23. An isolated LMP nucleic acid comprising a polynucleotide sequence encoding a polypeptide that functions as a modulator of a seed storage compound in a plant, wherein the

polynucleotide is selected from the group encodes a polypeptide selected from the group consisting of the sequences shown in Figure 1C, Figure 2C, Figure 3C, Figure 4C, and Figure 5C.

24. An isolated nucleic acid comprising a polynucleotide of least 60 consecutive nucleotides the LMP nucleic acid of claims 22 or 23.

25. An isolated nucleic acid comprising a polynucleotide having at least 70% sequence identity with the LMP nucleic acid of claims 22 or 23.

26. An isolated nucleic acid comprising a polynucleotide having at least 90% sequence identity with the LMP nucleic acid of claims 22 or 23.

27. An isolated nucleic acid comprising a polynucleotide complementary to the LMP nucleic acid of claims 22 or 23.

28. An isolated nucleic acid that hybridizes under stringent conditions to a second nucleic acid nucleic acid of claims 22 or 23.

**Figure 1A: Nucleotide sequence of PpCk63 (SEQ ID NO.:1)**

GGCGTCGTCTTCGATGGAACCCCGCGTCGGCAACAAGTATCGCCTTGGCCGGAAA  
ATTGGGAGTGGTTCCTTTGGTGAGATCTACCTGGGGACCAATCTCGTGACTCATGA  
GGAGGTCGGCATCAAGCTGGAGAGCATCAAGGCCAAGCATCCACAATTGCTTTAT  
GAGTCCAAGTTGTACCGTATTCTTCAAGGAGGAACTGGGATTCCCAACATCAGATG  
GTACGGAATTGAAGGAGACTATAATGTGATGGTTCTTGATCTTCTGGGACCCAGTC  
TTGAAGATCTTTTCAATTTCTGCAGCCGGAAATTCTCTTTGAAGACAGTTCTCATGC  
TTGCCGACCAGCTGATCAATCGAGTGGAGTATGTGCATGCCAAGAGTTTCCTCCAC  
AGGGACATAAAGCCTGACAATTTCTTGATGGGGCTAGGCAGGCGAGCAAATCAGG  
TCTATATGATTGACTTTGGTCTTGCAAAGAAGTATCGCGATCCCACTACTCATCAGC  
ACATTCCTTATAGAGAGAACAAAATCTTACTGGAACCGCTCGATATGCAAGTATCA  
ACACTCATCTTGGTATTGAACAAAGCAGGAGAGATGATCTGGAGTCTCTTGGATAT  
GTTCTCATGTATTTCTTGAGAGGCAGCCTGCCTTGGCAAGGAATGAAAGCAGGAAC  
CAAGAAGCAGAAGTATGAAAAAATCAGTGAGAAAAAGATGTCCACCCCTATAGAGT  
TCCTTTGTAAAGCTTACCCGTCTGAGTTTGCTTCATACTTCCACTACTGTGCGTCTC  
TTCGGTTCGATGACAAACCGGACTATGCTTACCTGAAGAGAATTTTCCGAGATCTCT  
TCATTCGTGAGGGTTTTTCAGTTTGATTATGTTTTCGACTGGACGATTTTGAAGTATC  
AGCAAACACATTTTTCTGGTGGTCCTCTCCGTCCAGCGGCTGCGGCGGGAGGTTT  
AAGTGGAGCAGCAGCAGCAGCGGCAGCAGGAATTGGTACAGTCCCAAGAGACGC  
CCAGCGAGCAATTGAGCCTACTGATGTTGCCGCTCGAACTCGAATGGTTGGTGGC  
ACTCGCTCTAGTGGATTAATCCACTGGACGCGTCAAAGCATAAGAGTACTAGCCC  
AGATGAAGCCGCTTCTAAGGACATAGCCCTTAGCGGTCTTGCAGAACCAGAGCGC  
ACGCATGCTTCTTCGTTTGTGCGGGGGAGCTCATCATCAAGGAGAGCTGTTGTTGG  
ATGTGCTAGGCCAGCAGGGTCAACAGAGGCGGGAGATGGAACGCGGGTGTGGC  
TGGCAAAATGGGCCCCACTAGCCTGCGCACATCAGCAGGAATGCAGAGGAGCTCT  
CCGGTGGCATCTACGGATCCCAAGCGGACGGGACGAGATTCTTATGCTGGAACT  
CCGGAAGAAATCCTAGTTCCTCTCGAAATTCGAAAGAGTGAGCACATTGGTTGAAC  
TGGGTCTGCATCTTGTTTCGAAGAGCATTACAACCTGTATCTGGCCTTGGTATCTGC  
TGTGGTTTAGGAATTTGGCCTTGTACTTGATTTGAAGAACAGGTTTCGTAAGAAATTG  
ATCAAATTTCAATGTGCGTGGGCGTCCAGTTCAGGATATGGTTGGTGGCTTGTGATG  
GATATATTATCTGTTTCTATCTTTGAAGGTGTTGCCCCAGCCGTATAGTTTTTCATCT  
TTCATAGCTTGTAGTTGGCAAAGCCCATTGCCATCTGTCAATATTCAGAGTGTGGTT  
GAGGGAGCTGTTTAGCCTTCTAGATAAGAGATCTGGATTGCGTTCTGGATTGCTGC  
ACAGACCTTGAAGATTTGTGGCTGTTTCG

**Figure 1B: Nucleotide sequence of the open reading frame of PpCk63 (SEQ ID NO.:2)**

ATGGAACCCCGCGTCGGCAACAAGTATCGCCTTGGCCGGAAAATTGGGAGTGGTT  
CCTTTGGTGAGATCTACCTGGGGACCAATCTCGTGACTCATGAGGAGGTCGGCAT  
CAAGCTGGAGAGCATCAAGGCCAAGCATCCACAATTGCTTTATGAGTCCAAGTTGT  
ACCGTATTCTTCAAGGAGGAACTGGGATTCCCAACATCAGATGGTACGGAATTGAA  
GGAGACTATAATGTGATGGTTCTTGATCTTCTGGGACCCAGTCTTGAAGATCTTTT  
AATTTCTGCAGCCGGAAATTCTCTTTGAAGACAGTTCTCATGCTTGGCAGCCAGCT  
GATCAATCGAGTGGAGTATGTGCATGCCAAGAGTTTCCTCCACAGGGACATAAAGC  
CTGACAATTTCTTGATGGGGCTAGGCAGGCGAGCAAATCAGGTCTATATGATTGAC  
TTTGGTCTTGC AAAGAAGTATCGCGATCCCACTACTCATCAGCACATTCTTATAGA  
GAGAACAAAATCTTACTGGAACCGCTCGATATGCAAGTATCAACACTCATCTTGGT

ATTGAACAAAGCAGGAGAGATGATCTGGAGTCTCTTGGATATGTTCTCATGTATTTCTT  
 TTGAGAGGCAGCCTGCCTTGGCAAGGAATGAAAGCAGGAACCAAGAAGCAGAAGT  
 ATGAAAAAATCAGTGAGAAAAAGATGTCCACCCCTATAGAGTTCCTTTGTAAAGCTT  
 ACCCGTCTGAGTTTGCTTCATACTTCCACTACTGTCGGTCTCTTCGGTTCGATGACA  
 AACCGGACTATGCTTACCTGAAGAGAATTTTCCGAGATCTCTTCATTTCGTGAGGGTT  
 TTCAGTTTGATTATGTTTTCGACTGGACGATTTTGAAGTATCAGCAAACACATTTTTC  
 TGGTGGTCTCTCCGTCCAGCGGCTGCGGCGGGAGGTTCAAGTGGAGCAGCAGC  
 AGCAGCGGCAGCAGGAATTGGTACAGTCCCAAGAGACGCCAGCGAGCAATTGAG  
 CCTACTGATGTTGCCGCTCGAACTCGAATGGTTGGTGC GACTCGCTCTAGTGGATT  
 AAATCCACTGGACGCGTCAAAGCATAAGAGTACTAGCCCAGATGAAGCCGCTTCTA  
 AGGACATAGCCCTTAGCGGTCTTG CAGAACCAGAGCGCACGCATGCTTCTTCGTTT  
 GTGCGGGGGAGCTCATCATCAAGGAGAGCTGTTGTTGGATGTGCTAGGCCAGCAG  
 GGTCAACAGAGGCGGGAGATGGAACGCGGGTGTGGCTGGCAAATGGGCCCA  
 CTAGCCTGCGCACATCAGCAGGAATGCAGAGGAGCTCTCCGGTGGCATCTACGGA  
 TCCCAAGCGGACGGGACGAGATTCTTATGCTGAAACTCCGGAAGAAATCCTAGTT  
 CCTCTCGAAATTCGAAAGAGTGA

**Figure 1C: Amino acid sequence of the open reading frame of PpCk63 (SEQ ID NO.:3)**

MEPRVGNKYRLGRKIGSGSFGIYLGTLNLVTHEEVGIKLESIKAKHPQLLYESKLYRILQ  
 GGTGIPNIRWYGIEGDYNVMVLDLLGPSLEDLNFNCSRKFSLKTVLMLADQLINRVEYV  
 HAKSFLHRDIKPDNFLMGLGRRANQVYMFDFGLAKKYRDPPTHQHIPPYRENKNLTGTA  
 RYASINHLGIEQSRRDDLES LGYVLMYFLRGS LPWQGMKAGTKKQKYEKISEKKMST  
 PIEFLCKAYPSEFASYFH YCRSLRFDDKPDYAYLKRIFRDLFIREGFQFDYVFDWILKY  
 QQTHFSGGPLRPA AAAAGSSGAAAAAAGIGTVPRDAQRAIEPTDVAARTRMVGATR  
 SSSLNPLDASKHKSTSPDEAASKDIALSGLAEPERTHASSFVRGSSSSRRAVVGCARP  
 AGSTEAGDGTRVLAGKMGPTSLRTSAGMQRSSPVASTDPKRTGRDSYAGNSGRNPS  
 SSRNSKE

**Figure 2A: Nucleotide sequence of BnCk01 (SEQ ID NO.:4)**

TCGTCCTCCTCCTACGCATCTCTTGGACGCTTCTCTTATTGTGCTGCGCCATCTTC  
 ATCTCGCGTGCGCCGTCGGCGCTTTCTCTTAACGGGAGCAGCTTCGATCCATATCC  
 ACGCGAAGAAGAGATCTCTCTTCCCACCGTGATGTCGAAAGCTCGTGTGTACACCG  
 ACGTGAACGTGATCCGTCCCAAAGAGTATTGGGATTACGAGTCTCTCAATGTTGAG  
 TGGGGGGAGCAAGATGATTACGAGGTAGTGAGGAAAGTGGGGAGAGGCCAAATACA  
 GTGAAGTTTTCGAGGGGATTAACATGAACAGCAACGAGAAATGTGTTATCAAGATT  
 CTTAAGCCTGTCAAGAAGAAGAAGATTAGAAGAGAGATTAAGATACTTCAGAATCTC  
 TGTGGAGGCCCAAATATTGTGAAGCTGCTCGATGTTGTCAGAGACCAACACTCAA  
 AACCCCAAGCTTGATATTTGAGTATGTTAACAGCACTGACTTTAAGGTTCTGTATCC  
 TACTTTGACTGACTATGACATCCGATACTACATCTACGAGCTGCTGAAGGCATTGG  
 ACTACTGCCACTCGCAAGGTATAATGCACAGAGATGTCAAGCCACACAATGTCATG  
 ATCGACCATGAGCTGCGCAAACCTTCGCCTTATAGATTGGGGTCTCGCTGAGTTTTA  
 TCATCCTGGAAAAGAGTATAACGTCCGTGTGGCCTCCAGGTACTTCAAGGGACCC  
 GAACTTTTAGTGGATTTACAGGACTATGACTATTCCTTGGACATGTGGAGCCTCGG  
 TTGCATGTTTGCTGGGATGATATCCGCAAGGAACCTTTCTTCTATGGCCATGACAA  
 CCAGGATCAGCTTGTCAA AATTGCCAAGGTCCTTGGAAACCGATGAATTGAATGCAT  
 ATCTGAACAAGTATCAGTTAGA ACTTGACGCTCAATTAGAGGCACTTGTTGGGAGA

CATAGCAGGAAGCCTTGGTCCAAATTCATCAATGCTGACAATAGGCATTTGGTCTC  
 ACCTGAGGCGATTGATTTCTGGACAAGCTACTTCGGTATGATCATCAAGACAGGT  
 TAACTGCAAAGAAGCAATGGCTCATCCTTATTTTCGCACAAGTCAAGGCAGCAGAG  
 AGCAGCAGAATGAGAACTTAATAGGACGTCTCTGAGTGTGTATCTGAAACTGGTGG  
 ATTGTCCTCCCGATAAGTTGTTGTTATCTTTTAAAGAAGTGGTGTAAATGTTTAACTT  
 TAAAAGTTTATCCGAGTATTTTCATCTATTTTCAATCTTCTGTAGAACTTCTATTGTTT  
 AATCTTGATATTTATAAAAAGGAATCTAGTTTTTTTTTAAAAAAAAAAAAAAAAAAAA

**Figure 2B: Nucleotide sequence of the open reading frame of BnCh01 (SEQ ID NO.: 5)**

ATGTCGAAAGCTCGTGTGTACACCGACGTGAACGTGATCCGTCCCAAAGAGTATTG  
 GGATTACGAGTCTCTCAATGTTGAGTGGGGGGAGCAAGATGATTACGAGGTAGTG  
 AGGAAAGTGGGGAGAGGCAAATACAGTGAAGTTTTTCGAGGGGATTAACATGAACA  
 GCAACGAGAAATGTGTTATCAAGATTCTTAAGCCTGTCAAGAAGAAGAAGATTAGAA  
 GAGAGATTAAGATACTTCAGAATCTCTGTGGAGGCCCAAATATTGTGAAGCTGCTC  
 GATGTTGTCAGAGACCAACACTCAAAAACCCCAAGCTTGATATTTGAGTATGTTAAC  
 AGCACTGACTTTAAGGTTCTGTATCCTACTTTGACTGACTATGACATCCGATACTAC  
 ATCTACGAGCTGCTGAAGGCATTGGACTACTGCCACTCGCAAGGTATAATGCACAG  
 AGATGTCAAGCCACACAATGTCATGATCGACCATGAGCTGCGCAAACCTTCGCCTTA  
 TAGATTGGGGTCTCGCTGAGTTTTATCATCCTGGAAAAGAGTATAACGTCCGTGTG  
 GCCTCCAGGTA CTTC AAGGGACCCGAAC TTTAGTGGATTTACAGGACTATGACTA  
 TTCCTTGGACATGTGGAGCCTCGGTTGCATGTTTGCTGGGATGATATTCGCAAGG  
 AACCTTTCTTCTATGGCCATGACAACCAGGATCAGCTTGTCAA AATTGCCAAGGTC  
 CTTGGAACCGATGAATTGAATGCATATCTGAACAAGTATCAGTTAGAACTTGACGCT  
 CAATTAGAGGCACTTGTTGGGAGACATAGCAGGAAGCCTTGGTCCAAATTCATCAA  
 TGCTGACAATAGGCATTTGGTCTCACCTGAGGCGATTGATTTCTGGACAAGCTAC  
 TTCGGTATGATCATCAAGACAGGTTAACTGCAAAGAAGCAATGGCTCATCCTTATT  
 TCGCACAAGTCAAGGCAGCAGAGAGCAGCAGAATGAGAACTTAA

**Figure 2C: Amino acid sequence of the open reading frame of BnCh01 (SEQ ID NO.: 6)**

MSKARVYTDVNVIRPKEYWDYESLNVEWGEQDDYEVVRKVGRGKYSEVFEGINMNS  
 NEKCVIKILKPVKKKIRREIKILQNL CGGNIVKLLDVVRDQHSKTPSLIFEYVNSTDFKV  
 LYPTLTDYDIRYYIYELLKALDYCHSQGIMHRDVKPHNVMIDHELRLRLIDWGLAEFYH  
 PGKEYNVRVASRYFKGPELLVDLQDYDYS LDMWVSLGCMFAGMIFRKEPFFYGHNDQD  
 QLVKIAKVLGTDELNAYLNKYQLELDAQLEALVGRHSRKPWSKFINADNRHLVSPEAID  
 FLDKLLRYDHQDRLTAKEAMAHPYFAQVKAESSRMRT

**Figure 3A: Nucleotide sequence of GmCh02 (SEQ ID NO.: 7)**

TTTCGTTCCCTCTGCTCCTCCCTTTCCATTCCCTTCCCTCTCCCTCCGCTCCC  
 CCAATTTCCCCCAAAC TTTCTCCCTCCGCGCTTCGCGCGCGCCCTCTCGCTCGC  
 CCTCCGCGCGCGCCTCCGCCCCTGGACGCCATGGCCCAGAAAATCGGAAAGTCC  
 ACGCGCCGCCCCGGCTCGTCCCAAGGCCCGCGTCTACGCCGATGTCAACGTC  
 GTTCGCCCCAAGGAATATTGGGACTACGAGTCCCTCAACGTCCAGTGGGGGGAGC  
 AAGATGATTATGAGGTGGTGAGGAAGGTGGGTAGGGGGAAATACAGTGAGGTGTT  
 CGAAGGTGTTCACTGCACGGACGGTGAGAAATGTGTCATTAAGATCCTGAAACCCG

TGAAGAAGAAGAAGATCAAGAGGGAAATCAAATATTACAGAATCTTTGTGGAGGG  
 CCTAATATTGTACAGTTGCTTGACATTGTTAGAGACCAACAGTCGAAGACCCCAAGT  
 CTTATATTTGAATATGTCAATAACACTGATTTTAAAGTGCTCTATCCTACATTGTCAG  
 ACTACGAAATACGATATTATATTTATGAACTTCTGAAGGCACTGGATTATTGCCATTC  
 ACAAGGAATCATGCATCGTGATGTGAAACCACATAATGTAATGATTGATCATGAGCA  
 GCGTAAGCTTCGCCTTATAGATTGGGGGCTTGCAGAGTTCTATCATCCTGGGAAAG  
 AATATAATGTTTCGCGTGGCATCAAGATATTTCAAAGGTCCTGAGCTTCTTGTTGATC  
 TTCAAGACTATGATTACTCTCTAGATTTATGGAGTCTTGGTTGTATGTTTGCTGGGA  
 TGATATTCGTAAGGAGCCATTCTTTTATGGACATGATAACTATGATCAACTGGTCA  
 AAATAGCTAAGGTACTTGGGACAGACGGATTAAGTGCATATTTAGATAAGTATCGAA  
 TAGAGTTGGACCCACATCTTGCAGCACTTATCGGGAGGCATAGCCGCAAGCCATG  
 GGCAAAGTTCATTAATGTAGAAAATCATCACATGGCTGTTCTGAGGCTGTTGACTT  
 TGTTGACAAGTTATTACGCTATGATCACCAAGAACGTCCCACAGCAAAGAAGCCA  
 TGGCCCATCCATACTTCAATCCAGTTAGAAATGCAGAAAGTAGTAGAACTCGTGCA  
 CATTGAGTTGCTGGTGCTTGTGTGATATGGTGTTCGTTCTTGTGAGGGGGCCTTC  
 AACAGTCTGAACTGTTATTTCTAGTGGAACTTGCTTTTAGTAGATTCCCCTGTTCTG  
 TTGTGAACTCATTTCAATTAGTTCTTTATCTAATCGTTTTGTTTCTGGTTTTTAAAT  
 TATTTTTGGGTGGAAAATTCAGCTCTCAACAATTTTATTTGATTGTGTGCCAAAATTT  
 GGAATGAAATGTTGATTCAATTTAGTAAAATTATACGCTTGTGTTCTTTTATGTAA  
 TACTGTAAATCAAATTTTATGGTTCCCAAAAAAAAAAAAAAAAAAAAAA

**Figure 3B: Nucleotide sequence of the open reading frame of GmCk02 (SEQ ID NO.: 8)**

ATGGCCCAGAAAATCGGAAAGTCCACGCGCCGCCCGGCTCGTCCTCCAAGGCC  
 CGCGTCTACGCCGATGTCAACGTCGTTCCGCCCAAGGAATATTGGGACTACGAGT  
 CCCTCAACGTCCAGTGGGGGGAGCAAGATGATTATGAGGTGGTGAGGAAGGTGG  
 GTAGGGGGAAATACAGTGAGGTGTTTCAAGGTGTTCACTGCACGGACGGTGAGAA  
 ATGTGTCATTAAGATCCTGAAACCCGTGAAGAAGAAGAAGATCAAGAGGGAAATCA  
 AAATATTACAGAATCTTTGTGGAGGGCCTAATATTGTACAGTTGCTTGACATTGTTA  
 GAGACCAACAGTCGAAGACCCCAAGTCTTATATTTGAATATGTCAATAACACTGATT  
 TAAAGTGCTCTATCCTACATTGTCAGACTACGAAATACGATATTATATTTATGAACT  
 TCTGAAGGCACTGGATTATTGCCATTCACAAGGAATCATGCATCGTGATGTGAAAC  
 CACATAATGTAATGATTGATCATGAGCAGCGTAAGCTTCGCCTTATAGATTGGGGG  
 CTTGCAGAGTTCTATCATCCTGGGAAAGAATATAATGTTTCGCGTGGCATCAAGATAT  
 TTCAAAGGTCCTGAGCTTCTTGTGATCTTCAAGACTATGATTACTCTCTAGATTTAT  
 GGAGTCTTGGTTGTATGTTTGCTGGGATGATATTCGTAAGGAGCCATTCTTTTATG  
 GACATGATAACTATGATCAACTGGTCAAATAGCTAAGGTACTTGGGACAGACGGA  
 TTAAGTGCATATTTAGATAAGTATCGAATAGAGTTGGACCCACATCTTGCAGCACTT  
 ATCGGGAGGCATAGCCGCAAGCCATGGGCAAAGTTCATTAATGTAGAAAATCATCA  
 CATGGCTGTTCTGAGGCTGTTGACTTTGTTGACAAGTTATTACGCTATGATCACCA  
 AGAACGTCCCACAGCAAAGAAGCCATGGCCATCCATACTTCAATCCAGTTAGAA  
 ATGCAGAAAGTAGTAGAACTCGTGACAT

**Figure 3C: Amino acid sequence of the open reading frame of GmCk02 (SEQ ID NO.:9)**

MAQKIGKSTRPGRSSSKARVYADVNVVRPKEYWDYESLNVQWGEQDDYEVVRKVGR  
 GKYSEVFEGVHCTDGEKCVIKILKPVKKKIKREIKILQNLCCGPNIVQLLDIVRDQQSKT

PSLIFEYVNNTDFKVL YPTLSDYEIRYYIYELLKALDYCHSQGIMHRDVKPHNV MIDHEQ  
 RKLRLIDWGLAEFYHPGKEYNVRVASRYFKGPELLVDLQDYDYSLDLWLSLGC MFAGMI  
 FRKEPFFYGHNDYDQLVKIAKVLGTDGLSAYLDKYRIELDPHLAALIGRHSRKPWAKFIN  
 VENHHMAVPEAVDFVDKLLRYDHQERPTAKEAMAHYPFNPVRNAESSRTRAH

**Figure 4A: Nucleotide sequence of ZmCk01 (SEQ ID NO.:10)**

CGACGATTTTCGTTCGATCGGTCCATCGTTCGTTCGGCTCGGTAGCCGCCGGATCGGG  
 ATCGGCCCGGCCCATGTCCAAGGCCAGGGTCTACACCGACGTCAACGTGCTGCG  
 CCCC AAGGAGTACTGGGACTACGAGGCGCTCACCGTTCAATGGGGTGAGCAGGAT  
 AACTATGAAGTTGTTAGGAAAGTTGGGAGAGGGAAATACAGTGAAGTCTTTGAGGG  
 TATTAATGTTAACAATAATGAGAAATGCATCATTAAAGATCCTCAAGCCTGTGAAGAA  
 AAAGAAGATCAAAAGAGAAAATTAATAATTCTTCAGAATCTTTATGGAGGCCCTAACAT  
 TGTGAAGCTGCTTGATATTGTTAGAGGTCAGCATTCAAAAACACCAAGCTTGATCTT  
 TGAGTATGTCAACAACACAGATTTCAAAGTGTGTACCCGACACTAACAGATTATGA  
 CATTTCGATATTACATCTATGAGTTACTCAAGGCATTAGATTACTGCCATTACAAGG  
 TATTATGCACCGAGATGTGAAGCCCCACAATGTTATGATAGATCATGAGCTTCGGA  
 AACTCCGGCTGATAGACTGGGGCCTTGCTGAATTCTATCATCCTGGCAAGGAATAT  
 AATGTCCGTGTTGCCTCCAGGTACTTCAAGGGTCTGAGCTTCTTGTTGACTTGCA  
 AGATTATGATTACTCTCTGGACATGTGGAGTCTGGGCTGCATGTTTGCTGGAATGA  
 TTTTTCGTAAGGAGTCATTTTTTTACGGTCATGACAACCATGATCAACTTGTTAAGAT  
 AGCCAAGTACTTGGAACAGATGGGCTGAATGTCTATTTGAACAAGTACAGAATTG  
 AGCTTGACCCTCAGCCC GAAGCCCTTGTTGGAAGGCATAGCAGGAAACCCTGGAC  
 AAAGTTTATTAATGCAGACAACCAACACCTAGTGTCTCATGAGGCCATAGATTTCT  
 CGACAAGCTTTTGC GTTATGATCATCAGGACAGGCTTACTGCACGTGAAGCTATGA  
 CACACCCATACTTCCAACAAGTGAGGGCTGCAGAGAACAGCAGGACGCGGGCATG  
 AGAGGCCAAAGACCAAAGACTGGTGCTAACCCCTGGATCAGATCATCTGCTCTTGGT  
 GGAGCCATGCTCTTACCAGTGACTTGCTTCAGTCGTATGTGATGAATAGTGGTTTG  
 GGGGTTTACTTGGCAAAAACAAACAGTAAAAAAAAAACCGACACGCAG

**Figure 4B: Nucleotide sequence of the open reading frame of ZmCk01 (SEQ ID NO.:11)**

ATGTCCAAGGCCAGGGTCTACACCGACGTCAACGTGCTGCGCCCCAAGGAGTACT  
 GGGACTACGAGGCGCTCACCGTTCAATGGGGTGAGCAGGATAACTATGAAGTTGT  
 TAGGAAAGTTGGGAGAGGGAAATACAGTGAAGTCTTTGAGGGTATTAATGTTAACA  
 ATAATGAGAAATGCATCATTAAAGATCCTCAAGCCTGTGAAGAAAAAGAAGATCAAAA  
 GAGAAATTAATAATTCTTCAGAATCTTTATGGAGGCCCTAACATTGTGAAGCTGCTTG  
 ATATTGTTAGAGGTCAGCATTCAAAAACACCAAGCTTGATCTTTGAGTATGTCAACA  
 ACACAGATTTCAAAGTGTGTACCCGACACTAACAGATTATGACATTTCGATATTACA  
 TCTATGAGTTACTCAAGGCATTAGATTACTGCCATTACAAGGTATTATGCACCGAG  
 ATGTGAAGCCCCACAATGTTATGATAGATCATGAGCTTCGGAAACTCCGGCTGATA  
 GACTGGGGCCTTGCTGAATTCTATCATCCTGGCAAGGAATATAATGTCCGTGTTGC  
 CTCCAGGTA CTTCAAGGGTCTGAGCTTCTTGTTGACTTGCAAGATTATGATTACTC  
 TCTGGACATGTGGAGTCTGGGCTGCATGTTTGCTGGAATGATTTTTCGTAAGGAGT  
 CATTTTTTTACGGTCATGACAACCATGATCAACTTGTTAAGATAGCCAAGGTA CTTG  
 GAACAGATGGGCTGAATGTCTATTTGAACAAGTACAGAATTGAGCTTGACCCTCAG  
 CCCGAAGCCCTTGTTGGAAGGCATAGCAGGAAACCCTGGACAAAGTTTATTAATGC  
 AGACAACCAACACCTAGTGTCTCATGAGGCCATAGATTTCTCGACAAGCTTTTGC

GTTATGATCATCAGGACAGGCTTACTGCACGTGAAGCTATGACACACCCATACTTC  
CAACAAGTGAGGGCTGCAGAGAACAGCAGGACGCGGGCATGA

**Figure 4C: Amino acid sequence of the open reading frame of ZmCk01 (SEQ ID NO.:12)**

MSKARVYTDVNVLRPKEYWDYEALTVQWGEQDNYEVVRKVGRGKYSEVFEGINVNN  
NEKCIKILKPVKKKIKREIKILQNLVGGPNIVKLLDIVRGQHSKTPSLIFEYVNNDFKVL  
YPTLTDYDIRYYIYELLKALDYCHSQGIMHRDVKPHNVIMIDHELRLRLIDWGLAEFYHP  
GKEYNVRVASRYFKGPELLVDLQDYDYSLDMWSLGCMFAGMIFRKESFFYGHNDHQ  
LVKIAKVLGTDGLNVYLNKYRIELDPQPEALVGRHSRKPWTKFINADNQHLVSHEAIDFL  
DKLLRYDHQDRLTAREAMTHPYFQQVRAAENSRTA

**Figure 5A: Nucleotide sequence of OsCk01 (SEQ ID NO.:13)**

TTTCGTTCTTGTCCCCTTCCCTCGTCGAAACCCTAGCTAGGCTCGCCGCCGGCGAC  
GCCCGCCTCGCCGATGCGCTCCCCACCTGACCGCTCCGCCTCGCTCTCCCGGG  
TCGTGTTGCCTCACCGCTGTGCCCGCCACCCCCACCTGGGCAGCGGGCGGGA  
TCCAGTCCCAGATCGGCGCCATGTCCAAGGCCAGGGTCTACACCGACGTCAACGT  
GCTGCGCCCCAAGGAGTACTGGGACTACGAGGCCCTCACCGTCCAATGGGGTGA  
GCAGGATGATTATGAAGTTGTCAGGAAAGTTGGGAGAGGAAAATACAGCGAGGTC  
TTTGAGGGTATTAATGTCAACAACAATGAAAAATGCATTATCAAATTCTCAAGCCT  
GTGAAGAAAAGAAGATCAAAAGAGAGATAAAAATACTTCAGAATCTTTGTGGAGGT  
CCAAATATCGTGAAACTGCTTGATATTGTTAGAGATCAGCACTCAAAAACACCAAGC  
TTGATATTTGAATATGTCAACAACACAGATTTCAAAGTGCTATACCCAACCCTGACA  
GATTATGACATCCGTTATTATATCTACGAGCTACTGAAGGCTTTAGATTACTGCCAT  
TCACAAGGCATCATGCACCGCGATGTGAAGCCTCACAATGTTATGATAGATCACGA  
GCTTCGAAAACCTCCGATTAATAGACTGGGGCCTTGCTGAATTCTACCACCCTGGCA  
AGGAATATAATGTCCGTGTTGCCTCAAGGTAAGGACTTCAAGGGACCTGAGCTTCTTGT  
GACTTGCAAGATTATGACTACTCTTTGGACATGTGGAGTCTTGGTTGCATGTTTGT  
GGAATGATCTTCCGCAAGGAGCCATTTTTCTATGGTCATGACAACCATGATCAACTT  
GTTAAGATCGCAAAGGTAAGGACTTGGGACAGACAGCCTGAATTCATACTTGAACAAGTA  
CCGTATTGAGCTTGACCCACAGCTTGAAGCCCTCGTTGGAAGGCACAGCAGGAAA  
CCTTGGTCAAAGTTTATTAATGCTGACAACCAACATCTAGTATCCCCAGAGGCCATC  
GATTTCTTGACAAGCTTCTTCGCTATGATCACCAGGATAGACTTACTGCACGTGAA  
GCTATGGCACACCCATACTTTCTCCAAGTGAGGGCTGCGGAGAACAGTAGGCCTC  
GTGCACAATAGACAAAAGAACGGGACTATTGCTGACTGTCGATTAGCAGCAGTCAT  
GGAGCTGCACCGTATTAGTACTTAGTCCAGTCACTTGTATTGGGAGTTTGGAGATT  
CGGTAATGGTTGAATCACCTGGCAAGAGGATTCTGGGAGAGCATGTAACAGCTCAT  
GTGTGCTTAACCGCAAACGTGGATTGTGATTTACTCGTGCTGTTTGTATCTTTGTT  
CTGACGATTGATTGAAACACATTGTTCTCGTTGAGTATAAATGTTTTTTTTTAATTTG  
CGAGTTGGTCCGTTTTAAA  
AAAAAAAAAAAAAAAAAAAAAAAAAAGGAAAAAAAAAAAAAAAAA

**Figure 5B: Nucleotide sequence of the open reading frame of OsCk01 (SEQ ID NO.:14)**

ATGTCCAAGGCCAGGGTCTACACCGACGTCAACGTGCTGCGCCCCAAGGAGTACT  
GGGACTACGAGGCCCTCACCGTCCAATGGGGTGAGCAGGATGATTATGAAGTTGT  
CAGGAAAGTTGGGAGAGGAAAATACAGCGAGGTCTTTGAGGGTATTAATGTCAACA  
ACAATGAAAAATGCATTATCAAAATTCTCAAGCCTGTGAAGAAAAAGAAGATCAAAA  
GAGAGATAAAAATACTTCAGAATCTTTGTGGAGGTCCAAATATCGTGAAACTGCTTG  
ATATTGTTAGAGATCAGCACTCAAAAACACCAAGCTTGATATTTGAATATGTCAACA  
ACACAGATTTCAAAGTGCTATACCCAACCCTGACAGATTATGACATCCGTTATTATA  
TCTACGAGCTACTGAAGGCTTTAGATTACTGCCATTCACAAGGCATCATGCACCGC  
GATGTGAAGCCTCACAAATGTTATGATAGATCACGAGCTTCGAAAACCTCCGATTAATA  
GACTGGGGCCTTGCTGAATTCTACCACCCTGGCAAGGAATATAATGTCCGTGTTGC  
CTCAAGGTACTTCAAGGGACCTGAGCTTCTTGTTGACTTGCAAGATTATGACTACTC  
TTTGGACATGTGGAGTCTTGTTGCATGTTTGCTGGAATGATCTTCCGCAAGGAGC  
CATTTTTCTATGGTCATGACAACCATGATCAACTTGTTAAGATCGCAAAGGTACTTG  
GGACAGACAGCCTGAATTCATACTTGAACAAGTACCGTATTGAGCTTGACCCACAG  
CTTGAAGCCCTCGTTGGAAGGCACAGCAGGAAACCTTGGTCAAAGTTTATTAATGC  
TGACAACCAACATCTAGTATCCCCAGAGGCCATCGATTTCTTGACAAGCTTCTTC  
GCTATGATCACCAGGATAGACTTACTGCACGTGAAGCTATGGCACACCCATACTTT  
CTCCAAGTGAGGGCTGCGGAGAACAGTAGGCCTCGTGCACAATAG

**Figure 5C: Amino acid sequence of the open reading frame of OsCk01 (SEQ ID NO.:15)**

MSKARVYTDVNVLRPKEYWDYEALTVQWGEQDDYEVVRKVGGRGKYSEVFEGINVNN  
NEKCIKILKPVKKKIKREIKILQNLCCGPNIVKLLDIVRDQHSKTPSLIFEYVNNTDFKVL  
YPTLTDYDIRYYIYELLKALDYCHSQGIMHRDVKPHNVMIDHELRLRLIDWGLAEFYHP  
GKEYNVRVASRYFKGPELLVDLQDYDYSLDMWSLGC MFAGMIFRKEPFFYGHNDHDQ  
LVKIAKVLGTDSLNSYLNKYRIELDPQLEALVGRHSRKPWSKFINADNQHLVSPEAIDFL  
DKLLRYDHQDRLTAREAMAHPYFLQVRAAENS RPRAQ