

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
25 June 2009 (25.06.2009)

PCT

(10) International Publication Number  
**WO 2009/077546 A2**

(51) International Patent Classification:  
C12N 15/82 (2006.01)

(21) International Application Number:  
PCT/EP2008/067679

(22) International Filing Date:  
17 December 2008 (17.12.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
07150061.5 17 December 2007 (17.12.2007) EP

(71) Applicant (for all designated States except US): **BASF Plant Science GmbH** [DE/DE]; 67056 Ludwigshafen (DE).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **ZANK, Thorsten** [DE/DE]; Seckenheimer Str. 4-6, 68165 Mannheim (DE).

(74) Agent: **POPP, Andreas**; GVX/B - C 6, Carl-Bosch-Str., 67056 Ludwigshafen (DE).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

- without international search report and to be republished upon receipt of that report
- with sequence listing part of description published separately in electronic form and available upon request from the International Bureau



**WO 2009/077546 A2**

(54) Title: LIPID METABOLISM PROTEIN AND USES THEREOF II (PHOSPHATE TRANSPORTER)

(57) Abstract: The present invention relates to phosphate transporter encoding polynucleotides and polypeptides. Further, the present invention relates to the use of the aforementioned nucleic acid sequences and proteins in transgenic plants. In particular, the invention is directed to methods for manipulating sugar-related compounds and for increasing oil level and altering the fatty acid composition in plants and seeds. The invention further relates to methods of using these novel plant polypeptides to stimulate plant growth and/or to increase yield and/or composition of seed storage compounds.

## LIPID METABOLISM PROTEIN AND USES THEREOF II (PHOSPHATE TRANSPORTER)

Described herein are inventions in the field of genetic engineering of plants, including iso-  
5 lated nucleic acid molecules encoding a phosphate transporter to improve agronomic, horti-  
cultural, and quality traits. This invention relates generally to nucleic acid sequences en-  
coding proteins that are related to the presence of seed storage compounds in plants.  
More specifically, the present invention relates to a phosphate transporter nucleic acid se-  
10 quences to be applied as sugar and lipid metabolism regulators and the use of these se-  
quences in transgenic plants. In particular, the invention is directed to methods for manipu-  
lating sugar-related compounds and for increasing oil level and altering the fatty acid com-  
position in plants and seeds. The invention further relates to methods of using these novel  
plant polypeptides to stimulate plant growth and/or to increase yield and/or composition of  
seed storage compounds.

15  
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 fatty acid content  
20 or altered fatty acid composition. 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 bio-  
chemists to affect plant metabolism giving rise to the production of specific higher-value  
25 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, Sci-  
ence 268:681-686), and non-traditional oil seed plants such as tobacco (Cahoon et al.  
1992, Proc. Natl. Acad. Sci. USA 89:11184-11188).

30  
Plant seed oils comprise both neutral and polar lipids (see Table 1). The neutral lipids con-  
sists primarily of triacylglycerol, which is the main storage lipid that accumulates in oil bod-  
ies in seeds. The polar lipids are mainly found in the various membranes of the seed cells,  
e.g. the microsomal membranes, the cell membrane and the mitochondrial and plastidial  
35 membranes. 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  
5 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).

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.  
10 235:25-31; 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 (AC-Case). 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  
15 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 de-  
20 saturase 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,  
25 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).

30 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.  
35 During the storage phase of seeds, sucrose is converted in the cytosol to provide the meta-

bolic 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, phosphoenolpyruvate 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.

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).

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.

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.

10 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; Beau-  
15 doin 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.

20 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 which have the capacity to confer altered or increased oil production to its host plant and to other plant species.

Thus, the technical problem underlying the present invention may be seen as the provision of means and methods for complying with the aforementioned needs. The technical problem is solved by the embodiments characterized in the claims and herein below. In principle, this invention discloses nucleic acid sequences from *Arabidopsis thaliana*. 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 canola, linseed, soybean, sunflower, maize, oat, rye, barley, wheat, rice, pepper, tagetes, cotton, oil palm, coconut palm, flax, castor, peanut, cranmbe, and *Jatropha*, which are oilseed plants  
35 containing high amounts of lipid compounds.

Specifically, the present invention relates to a polynucleotide comprising a nucleic acid sequences selected from the group consisting of:

- 5 (a) a nucleic acid sequence as shown in SEQ ID NO: 1;  
(b) a nucleic acid sequence encoding a polypeptide having an amino acid sequence as shown in SEQ ID NO: 2;  
(c) a nucleic acid sequence which is at least 70% identical to the nucleic acid sequence of (a) or (b), wherein said nucleic acid sequence encodes a polypeptide or biologically active portion thereof having phosphate transporter activity ; and  
10 (d) a nucleic acid sequence being a fragment of any one of (a) to (c), wherein said fragment encodes a polypeptide or biologically active portion thereof having phosphate transporter activity.

15 The term "polynucleotide" as used in accordance with the present invention relates to a polynucleotide comprising a nucleic acid sequence which encodes a polypeptide being a transporter protein, i.e. a polypeptide capable of transporting ions and, preferably, Pi, between different cellular compartments. More preferably, the said transporter protein is a phosphate transporter. The phosphate transporter polypeptides encoded by the polynucleo-  
20 tides of the present invention shall be capable of increasing the amount of seed storage compounds, preferably, fatty acids or lipids, when present in plant seeds. The polypeptides encoded by the polynucleotide of the present invention are also referred to as lipid metabolism proteins (LMP) herein below. Suitable assays for measuring the activities mentioned before are described in the accompanying Examples.

25 Preferably, the polynucleotide of the present invention upon expression in the seed of a transgenic plant is capable of significantly increasing the amount by weight of at least one seed storage compound. More preferably, such an increase as referred to in accordance with the present invention is an increase of the amount by weight of at least 1, 2.5, 5, 7.5,  
30 10, 12.5, 15, 17.5, 20, 22.5 or 25 % as compared to a control. Whether an increase is significant can be determined by statistical tests well known in the art including, e.g., Student's t-test. The percent increase rates of a seed storage compound are, preferably, determined compared to an empty vector control. An empty vector control is a transgenic plant, which has been transformed with the same vector or construct as a transgenic plant according to  
35 the present invention except for such a vector or construct is lacking the polynucleotide of

the present invention. Alternatively, an untreated plant (i.e. a plant which has not been genetically manipulated) or a wildtype regenerate from the in vitro culture may be used as a control.

5 A polynucleotide encoding a polypeptide having a biological activity as specified above has been obtained in accordance with the present invention from *Arabidopsis thaliana*. The corresponding polynucleotides, preferably, comprises the nucleic acid sequence shown in SEQ ID NO: 1 encoding a polypeptide having the amino acid sequence of SEQ ID NO: 2. It is to be understood that a polypeptide having an amino acid sequence as shown in SEQ ID NO:  
10 2 may be also encoded due to the degenerated genetic code by other polynucleotides as well.

Moreover, the term "polynucleotide" as used in accordance with the present invention further encompasses variants of the aforementioned specific polynucleotides. Said variants  
15 may represent orthologs, paralogs or other homologs of the polynucleotide of the present invention.

The polynucleotide variants, preferably, also comprise a nucleic acid sequence characterized in that the sequence can be derived from the aforementioned specific nucleic acid sequences shown in SEQ ID NO: 1 by at least one nucleotide substitution, addition and/or  
20 deletion whereby the variant nucleic acid sequence shall still encode a polypeptide having a biological activity as specified above. Variants also encompass polynucleotides comprising a nucleic acid sequence which is capable of hybridizing to the aforementioned specific nucleic acid sequences, preferably, under stringent hybridization conditions. These stringent  
25 conditions are known to the skilled worker and can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N. Y. (1989), 6.3.1-6.3.6. A preferred example for stringent hybridization conditions are hybridization conditions in  $6 \times$  sodium chloride/sodium citrate (= SSC) at approximately  $45^{\circ}\text{C}$ , followed by one or more wash steps in  $0.2 \times$  SSC, 0.1% SDS at  $50$  to  $65^{\circ}\text{C}$ . The skilled worker knows that these hybridization conditions differ depending  
30 on the type of nucleic acid and, for example when organic solvents are present, with regard to the temperature and concentration of the buffer. For example, under "standard hybridization conditions" the temperature differs depending on the type of nucleic acid between  $42^{\circ}\text{C}$  and  $58^{\circ}\text{C}$  in aqueous buffer with a concentration of  $0.1$  to  $5 \times$  SSC (pH 7.2). If organic solvent is present in the abovementioned buffer, for example 50% formamide, the temperature  
35 under standard conditions is approximately  $42^{\circ}\text{C}$ . The hybridization conditions for

DNA:DNA hybrids are, preferably,  $0.1 \times$  SSC and  $20^{\circ}\text{C}$  to  $45^{\circ}\text{C}$ , preferably between  $30^{\circ}\text{C}$  and  $45^{\circ}\text{C}$ . The hybridization conditions for DNA:RNA hybrids are, preferably,  $0.1 \times$  SSC and  $30^{\circ}\text{C}$  to  $55^{\circ}\text{C}$ , preferably between  $45^{\circ}\text{C}$  and  $55^{\circ}\text{C}$ . The abovementioned hybridization temperatures are determined for example for a nucleic acid with approximately 100 bp (= base pairs) in length and a G + C content of 50% in the absence of formamide. The skilled worker knows how to determine the hybridization conditions required by referring to textbooks such as the textbook mentioned above, or the following textbooks: Sambrook et al., "Molecular Cloning", Cold Spring Harbor Laboratory, 1989; Hames and Higgins (Ed.) 1985, "Nucleic Acids Hybridization: A Practical Approach", IRL Press at Oxford University Press, Oxford; Brown (Ed.) 1991, "Essential Molecular Biology: A Practical Approach", IRL Press at Oxford University Press, Oxford. Alternatively, polynucleotide variants are obtainable by PCR-based techniques such as mixed oligonucleotide primer-based amplification of DNA, i.e. using degenerated primers against conserved domains of the polypeptides of the present invention. Conserved domains of the polypeptide of the present invention may be identified by a sequence comparison of the nucleic acid sequences of the polynucleotides or the amino acid sequences of the polypeptides of the present invention. Oligonucleotides suitable as PCR primers as well as suitable PCR conditions are described in the accompanying Examples. As a template, DNA or cDNA from bacteria, fungi, plants or animals may be used. Further, variants include polynucleotides comprising nucleic acid sequences which are at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98% or at least 99% identical to the nucleic acid sequences shown in SEQ ID NO: 1 retaining a biological activity as specified above. Moreover, also encompassed are polynucleotides which comprise nucleic acid sequences encoding amino acid sequences which are at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98% or at least 99% identical to the amino acid sequences shown in SEQ ID NO: 2 wherein the polypeptide comprising the amino acid sequence retains a biological activity as specified above. The percent identity values are, preferably, calculated over the entire amino acid or nucleic acid sequence region. A series of programs based on a variety of algorithms is available to the skilled worker for comparing different sequences. In this context, the algorithms of Needleman and Wunsch or Smith and Waterman give particularly reliable results. To carry out the sequence alignments, the program PileUp (J. Mol. Evolution., 25, 351-360, 1987, Higgins et al., CABIOS, 5 1989: 151-153) or the programs Gap and BestFit (Needleman and Wunsch (J. Mol. Biol. 48; 443-453 (1970)) and Smith and Waterman (Adv. Appl. Math. 2; 482-489 (1981))), which are part of the GCG software packet [Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA 53711 (1991)], are to be

used. The sequence identity values recited above in percent (%) are to be determined, preferably, using the program GAP over the entire sequence region with the following settings: Gap Weight: 50, Length Weight: 3, Average Match: 10.000 and Average Mismatch: 0.000, which, unless otherwise specified, shall always be used as standard settings for sequence alignments. For the purposes of the invention, the percent sequence identity between two nucleic acid or polypeptide sequences can be also determined using the Vector NTI 7.0 (PC) software package (InforMax, 7600 Wisconsin Ave., Bethesda, MD 20814). A gap-opening penalty of 15 and a gap extension penalty of 6.66 are used for determining the percent identity of two nucleic acids. A gap-opening penalty of 10 and a gap extension penalty of 0.1 are used for determining the percent identity of two polypeptides. All other parameters are set at the default settings. For purposes of a multiple alignment (Clustal W algorithm), the gap-opening penalty is 10, and the gap extension penalty is 0.05 with blosum62 matrix. It is to be understood that for the purposes of determining sequence identity when comparing a DNA sequence to an RNA sequence, a thymidine nucleotide sequence is equivalent to an uracil nucleotide.

A polynucleotide comprising a fragment of any of the aforementioned nucleic acid sequences is also encompassed as a polynucleotide of the present invention. The fragment shall encode a polypeptide which still has a biological activity as specified above. Accordingly, the polypeptide may comprise or consist of the domains of the polypeptide of the present invention conferring the said biological activity. A fragment as meant herein, preferably, comprises at least 20, at least 50, at least 100, at least 250 or at least 500 consecutive nucleotides of any one of the aforementioned nucleic acid sequences or encodes an amino acid sequence comprising at least 20, at least 30, at least 50, at least 80, at least 100 or at least 150 consecutive amino acids of any one of the aforementioned amino acid sequences.

The variant polynucleotides or fragments referred to above, preferably, encode polypeptides retaining at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80% or at least 90% of the phosphate transporter activity exhibited by the polypeptide shown in SEQ ID NO: 2. The activity may be tested as described in the accompanying Examples.

The polynucleotides of the present invention either essentially consist of the aforementioned nucleic acid sequences or comprise the aforementioned nucleic acid sequences.

Thus, they may contain further nucleic acid sequences as well. Preferably, the polynucleotide of the present invention may comprise in addition to an open reading frame further untranslated sequence at the 3' and at the 5' terminus of the coding gene region: at least 500, preferably 200, more preferably 100 nucleotides of the sequence upstream of the 5' terminus of the coding region and at least 100, preferably 50, more preferably 20 nucleotides of the sequence downstream of the 3' terminus of the coding gene region.

Furthermore, the polynucleotides of the present invention may encode fusion proteins wherein one partner of the fusion protein is a polypeptide being encoded by a nucleic acid sequence recited above. Such fusion proteins may comprise as additional part other enzymes of the fatty acid or lipid biosynthesis pathways, polypeptides for monitoring expression (e.g., green, yellow, blue or red fluorescent proteins, alkaline phosphatase and the like) or so called "tags" which may serve as a detectable marker or as an auxiliary measure for purification purposes. Tags for the different purposes are well known in the art and comprise FLAG-tags, 6-histidine-tags, MYC-tags and the like.

Variant polynucleotides as referred to in accordance with the present invention may be obtained by various natural as well as artificial sources. For example, polynucleotides may be obtained by in vitro and in vivo mutagenesis approaches using the above mentioned mentioned specific polynucleotides as a basis. Moreover, polynucleotids being homologs or orthologs may be obtained from various animal, plant, bacteria or fungus species. Paralogs may be identified from *Arabidopsis thaliana*.

The polynucleotide of the present invention shall be provided, preferably, either as an isolated polynucleotide (i.e. isolated from its natural context such as a gene locus) or in genetically modified or exogenously (i.e. artificially) manipulated form. An isolated polynucleotide can, for example, comprise less than approximately 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in the genomic DNA of the cell from which the nucleic acid is derived. The polynucleotide, preferably, is double or single stranded DNA including cDNA or RNA. The term encompasses single- as well as double- stranded polynucleotides. Moreover, comprised are also chemically modified polynucleotides including naturally occurring modified polynucleotides such as glycosylated or methylated polynucleotides or artificial modified ones such as biotinylated polynucleotides.

The polynucleotide encoding a polypeptide having a biological activity as specified encompassed by the present invention is also, preferably, a polynucleotide having a nucleic acid sequence which has been adapted to the specific codon-usage of the organism, e.g., the plant species, in which the polynucleotide shall be expressed (i.e. the target organism). This is, in general, achieved by changing the codons of a nucleic acid sequence obtained from a first organism (i.e. the donor organism) encoding a given amino acid sequence into the codons normally used by the target organism whereby the amino acid sequence is retained. It is in principle acknowledged that the genetic code is redundant (i.e. degenerated). Specifically, 61 codons are used to encode only 20 amino acids. Thus, a majority of the 20 amino acids will be encoded by more than one codon. The codons for the amino acids are well known in the art and are universal to all organisms. However, among the different codons which may be used to encode a given amino acid, each organism may preferably use certain codons. The presence of rarely used codons in a nucleic acid sequence will result a depletion of the respective tRNA pools and, thereby, lower the translation efficiency. Thus, it may be advantageous to provide a polynucleotide comprising a nucleic acid sequence encoding a polypeptide as referred to above wherein said nucleic acid sequence is optimized for expression in the target organism with respect to the codon usage. In order to optimize the codon usage for a target organism, a plurality of known genes from the said organism may be investigated for the most commonly used codons encoding the amino acids. In a subsequent step, the codons of a nucleic acid sequence from the donor organism will be optimized by replacing the codons in the donor sequence by the codons most commonly used by the target organism for encoding the same amino acids. It is to be understood that if the same codon is used preferably by both organisms, no replacement will be necessary. For various target organisms, tables with the preferred codon usages are already known in the art; see e.g., <http://www.kazusa.or.jp/Kodon/E.html>. Moreover, computer programs exist for the optimization, e.g., the Leto software, version 1.0 (Entelechon GmbH, Germany) or the GeneOptimizer (Geneart AG, Germany). For the optimization of a nucleic acid sequence, several criteria may be taken into account. For example, for a given amino acid, always the most commonly used codon may be selected for each codon to be exchanged. Alternatively, the codons used by the target organism may replace those in a donor sequence according to their naturally frequency. Accordingly, at some positions even less commonly used codons of the target organism will appear in the optimized nucleic acid sequence. The distribution of the different replacement codons of the target organism to the donor nucleic acid sequence may be randomly. Preferred target organisms in accordance with the present invention are soybean or canola (Brassica) species. Preferably, the

polynucleotide of the present invention has an optimized nucleic acid for codon usage in the envisaged target organism wherein at least 20%, at least 40%, at least 60%, at least 80% or all of the relevant codons are adapted.

5 It has been found in the studies underlying the present invention that the polypeptides being encoded by the polynucleotides of the present invention is a phosphate transporter polypeptide involved in the regulation of seed storage compounds. Moreover, the polypeptides encoded by the polynucleotides of the present invention are, advantageously, capable of increasing the amount of seed storage compounds in plants significantly. Thus, the polynucleotides of the present invention are, in principle, useful for the enrichment and synthesis  
10 of seed storage compounds such as fatty acids or lipids. Moreover, they may be used to generate transgenic plants or seeds thereof having a modified, preferably increased, amount of seed storage compounds. Such transgenic plants or seeds may be used for the manufacture of seed oil or other lipid and/or fatty acid containing compositions.

15 Further, the present invention relates to a vector comprising the polynucleotide of the present invention. Preferably, the vector is an expression vector.

The term "vector", preferably, encompasses phage, plasmid, viral or retroviral vectors as well as artificial chromosomes, such as bacterial or yeast artificial chromosomes. Moreover, the term also relates to targeting constructs which allow for random or site-directed integration of the targeting construct into genomic DNA. Such target constructs, preferably, comprise DNA of sufficient length for either homologous recombination or heterologous insertion as described in detail below. The vector encompassing the polynucleotides of the present  
20 invention, preferably, further comprises selectable markers for propagation and/or selection in a host. The vector may be incorporated into a host cell by various techniques well known in the art. If introduced into a host cell, the vector may reside in the cytoplasm or may be incorporated into the genome. In the latter case, it is to be understood that the vector may further comprise nucleic acid sequences which allow for homologous recombination or heterologous insertion, see below. Vectors can be introduced into prokaryotic or eukaryotic  
25 cells via conventional transformation or transfection techniques. An "expression vector" according to the present invention is characterized in that it comprises an expression control sequence such as promoter and/or enhancer sequence operatively linked to the polynucleotide of the present invention Preferred vectors, expression vectors and transformation  
30 or transfection techniques are specified elsewhere in this specification in detail.

Furthermore, the present invention encompasses a host cell comprising the polynucleotide or vector of the present invention.

5 Host cells are primary cells or cell lines derived from multicellular organisms such as plants or animals. Furthermore, host cells encompass prokaryotic or eukaryotic single cell organisms (also referred to as microorganisms), e.g. bacteria or fungi including yeast or bacteria. Primary cells or cell lines to be used as host cells in accordance with the present invention may be derived from the multicellular organisms, preferably from plants. Specifically preferred host cells, microorganisms or multicellular organism from which host cells may be  
10 obtained are disclosed below.

The polynucleotides or vectors of the present invention may be incorporated into a host cell or a cell of a transgenic non-human organism by heterologous insertion or homologous recombination. "Heterologous" as used in the context of the present invention refers to a  
15 polynucleotide which is inserted (e.g., by ligation) or is manipulated to become inserted to a nucleic acid sequence context which does not naturally encompass the said polynucleotide, e.g., an artificial nucleic acid sequence in a genome of an organism. Thus, a heterologous polynucleotide is not endogenous to the cell into which it is introduced, but has been obtained from another cell. Generally, although not necessarily, such heterologous polynucleotides encode proteins that are normally not produced by the cell expressing the said heterologous polynucleotide. An expression control sequence as used in a targeting construct or expression vector is considered to be "heterologous" in relation to another sequence (e.g., encoding a marker sequence or an agronomically relevant trait) if said two  
20 sequences are either not combined or operatively linked in a different way in their natural environment. Preferably, said sequences are not operatively linked in their natural environment (i.e. originate from different genes). Most preferably, said regulatory sequence is covalently joined (i.e. ligated) and adjacent to a nucleic acid to which it is not adjacent in its natural environment. "Homologous" as used in accordance with the present invention relates to the insertion of a polynucleotide in the sequence context in which the said polynucleotide naturally occurs. Usually, a heterologous polynucleotide is also incorporated into a cell by homologous recombination. To this end, the heterologous polynucleotide is flanked by nucleic acid sequences being homologous to a target sequence in the genome of a host cell or a non-human organism. Homologous recombination now occurs between the homologous sequences. However, as a result of the homologous recombination of the flank-  
35

ing sequences, the heterologous polynucleotide will be inserted, too. How to prepare suitable target constructs for homologous recombination and how to carry out the said homologous recombination is well known in the art.

5 Also provided in accordance with the present invention is a method for the manufacture of a polypeptide having phosphate transporter activity comprising:

- (a) expressing the polynucleotide of the present invention in a host cell; and
- (b) obtaining the polypeptide encoded by said polynucleotide from the host cell.

10 The polypeptide may be obtained, for example, by all conventional purification techniques including affinity chromatography, size exclusion chromatography, high pressure liquid chromatography (HPLC) and precipitation techniques including antibody precipitation. It is to be understood that the method may – although preferred -not necessarily yield an essentially pure preparation of the polypeptide. It is to be understood that depending on the host  
15 cell which is used for the aforementioned method, the polypeptides produced thereby may become posttranslationally modified or processed otherwise.

The present invention, moreover, pertains to a polypeptide encoded by the polynucleotide of the present invention or which is obtainable by the aforementioned method of the present  
20 invention.

The term “polypeptide” as used herein encompasses essentially purified polypeptides or polypeptide preparations comprising other proteins in addition. Further, the term also relates to the fusion proteins or polypeptide fragments being at least partially encoded by the  
25 polynucleotide of the present invention referred to above. Moreover, it includes chemically modified polypeptides. Such modifications may be artificial modifications or naturally occurring modifications such as phosphorylation, glycosylation, myristylation and the like. The terms “polypeptide”, “peptide” or “protein” are used interchangeable throughout this specification. The polypeptide of the present invention shall exhibit the biological activities referred  
30 to above, i.e. it should be capable of transporting compounds such as ions (most preferably phosphate) and, more preferably, it shall be capable of increasing the amount of seed storage compounds, preferably, fatty acids or lipids, when present in plant seeds as referred to above. Most preferably, if present in plant seeds, the polypeptide shall be capable of significantly increasing the seed storage of lipids.

Encompassed by the present invention is, furthermore, an antibody which specifically recognizes the polypeptide of the invention.

Antibodies against the polypeptides of the invention can be prepared by well known methods using a purified polypeptide according to the invention or a suitable fragment derived therefrom as an antigen. A fragment which is suitable as an antigen may be identified by antigenicity determining algorithms well known in the art. Such fragments may be obtained either from the polypeptide of the invention by proteolytic digestion or may be a synthetic peptide. Preferably, the antibody of the present invention is a monoclonal antibody, a polyclonal antibody, a single chain antibody, a human or humanized antibody or primatized, chimerized or fragment thereof. Also comprised as antibodies by the present invention are: a bispecific antibody, a synthetic antibody, an antibody fragment, such as Fab, Fv or scFv fragments etc., or a chemically modified derivative of any of these. The antibody of the present invention shall specifically bind (i.e. does significantly not cross react with other polypeptides or peptides) to the polypeptide of the invention. Specific binding can be tested by various well known techniques. Antibodies or fragments thereof can be obtained by using methods which are described, e.g., in Harlow and Lane "Antibodies, A Laboratory Manual", CSH Press, Cold Spring Harbor, 1988. Monoclonal antibodies can be prepared by the techniques originally described in Köhler and Milstein, Nature 256 (1975), 495, and Galfré, Meth. Enzymol. 73 (1981), 3, which comprise the fusion of mouse myeloma cells to spleen cells derived from immunized mammals. The antibodies can be used, for example, for the immunoprecipitation, immunolocalization or purification (e.g., by affinity chromatography) of the polypeptides of the invention as well as for the monitoring of the presence of said variant polypeptides, for example, in recombinant organisms, and for the identification of compounds interacting with the proteins according to the invention.

The present invention also relates to a transgenic non-human organism comprising the polynucleotide, the vector or the host cell of the present invention. Preferably, said non-human transgenic organism is a plant.

The term "non-human transgenic organism", preferably, relates to a plant, an animal or a multicellular microorganism. The polynucleotide or vector may be present in the cytoplasm of the organism or may be incorporated into the genome either heterologous or by homologous recombination. Host cells, in particular those obtained from plants or animals, may be introduced into a developing embryo in order to obtain mosaic or chimeric organisms, i.e.

non-human transgenic organisms comprising the host cells of the present invention. Preferably, the non-human transgenic organism expresses the polynucleotide of the present invention in order to produce the polypeptide in an amount resulting in a detectable transporter activity due to the presence of the said polypeptide. Suitable transgenic organisms are, preferably, all those organisms which are capable of synthesizing fatty acids or lipids. Preferred organisms and methods for transgenesis are disclosed in detail below. A transgenic organism or tissue may comprise one or more transgenic cells. Preferably, the organism or tissue is substantially consisting of transgenic cells (i.e., more than 80%, preferably 90%, more preferably 95%, most preferably 99% of the cells in said organism or tissue are transgenic). The term "transgene" as used herein refers to any nucleic acid sequence, which is introduced into the genome of a cell or which has been manipulated by experimental manipulations including techniques such as chimera- or genoplasty. Preferably, said sequence is resulting in a genome which is significantly different from the overall genome of an organism (e.g., said sequence, if endogenous to said organism, is introduced into a location different from its natural location, or its copy number is increased or decreased). A transgene may comprise an endogenous polynucleotide (i.e. a polynucleotide having a nucleic acid sequence obtained from the same organism or host cell) or may be obtained from a different organism or host cell, wherein said different organism is, preferably an organism of another species and the said different host cell is, preferably, a different microorganism, a host cell of a different origin or derived from a an organism of a different species.

Particularly preferred as a plant to be used in accordance with the present invention are oil producing plant species. Most preferably, the said plant is 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,

The present invention relates to a method for the manufacture of a lipid and/or a fatty acid comprising the steps of:

- (a) cultivating the host cell or the transgenic non-human organism of the present invention under conditions allowing synthesis of the said lipid or fatty acid; and
- (b) obtaining the said lipid and/or fatty acid from the host cell or the transgenic non-human organism.

The term "lipid" and "fatty acid" as used herein refer, preferably, to those recited in Table 1 (for lipids) and Table 2 (for fatty acids), below. However, the terms, in principle, also en-

compass other lipids or fatty acids which can be obtained by the lipid metabolism in a host cell or an organism referred to in accordance with the present invention.

5 In a preferred embodiment of the aforementioned method of the present invention, the said lipid and/or fatty acids constitute seed oil.

Moreover, the present invention pertains to a method for the manufacture of a plant having a modified amount of a seed storage compound, preferably a lipid or a fatty acid, comprising the steps of:

- 10 (a) introducing the polynucleotide or the vector of the present invention into a plant cell; and
- (b) generating a transgenic plant from the said plant cell, wherein the polypeptide encoded by the polynucleotide modifies the amount of the said seed storage compound in the transgenic plant.

15 The term "seed storage compound" as used herein, preferably, refers to compounds being a sugar, a protein, or, more preferably, a lipid or a fatty acid. Preferably, the amount of said seed storage compound is significantly increased compared to a control, preferably an empty vector control as specified above. The increase is, more preferably, an increase in

20 the amount by weight of at least 1, 2.5, 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5 or 25 % as compared to a control.

It is to be understood that the polynucleotides or the vector referred to in accordance with the above method of the present invention may be introduced into the plant cell by any of

25 the aforementioned insertion or recombination techniques.

The aforementioned method of the present invention may be also used to manufacture a plant having altered total oil content in its seeds or a plant having altered total seed oil content and/or altered levels of seed storage compounds in its seeds. Such plants are suitable

30 sources for seed oil and may be used for the large scale manufacture thereof.

Further preferred embodiments of the compounds, methods and uses according to the present invention are described in the following. Moreover, the terms used above will be explained in more detail.

The present invention provides novel isolated nucleic acid and amino acid sequences, i.e., the polynucleotides and polypeptides of the present invention, associated with the metabolism of seed storage compounds in plants.

5 Preferably provided is a polynucleotide comprising a nucleic acid from *Arabidopsis thaliana* encoding the phosphate transporter polypeptide of the present invention, i.e. a Lipid Metabolism Protein (LMP), or a portion thereof. These sequences may be used to modify or increase lipids and fatty acids, cofactors and enzymes in microorganisms and plants.

10 *Arabidopsis* plants are known to produce considerable amounts of fatty acids like linoleic and linolenic acid (see, e.g., Table 2) and for their close similarity in many aspects (gene homology etc.) to the oil crop plant *Brassica*. Therefore, nucleic acid molecules originating from a plant like *Brassica napus* or related organisms including *Arabidopsis* (i.e. the polynucleotides of the present invention) are especially suited to modify the lipid and fatty acid  
15 metabolism in a host such as the host cells or transgenic non-human organisms of the present invention, especially in microorganisms and plants. Furthermore, nucleic acids from the plant *Arabidopsis thaliana* or related organisms can be used to identify those DNA sequences and enzymes in other species, which are useful to modify the biosynthesis of precursor molecules of fatty acids in the respective organisms.

20 The present invention further provides an isolated nucleic acid comprising a fragment of at least 15 nucleotides of a polynucleotide of the present invention, preferably, a polynucleotide comprising a nucleic acid from a plant encoding the polypeptides of the present invention.

25 The present invention, thus, also encompasses an oligonucleotide which specifically binds to the polynucleotides of the present invention. Binding as meant in this context refers to hybridization by Watson-Crick base pairing discussed elsewhere in the specification in detail. An oligonucleotide as used herein has a length of at most 100, at most 50, at most 40,  
30 at most 30 or at most 20 nucleotides in length which are complementary to the nucleic acid sequence of the polynucleotides of the present invention. The sequence of the oligonucleotide is, preferably, selected so that a perfect match by Watson-Crick base pairing will be obtained. The oligonucleotides of the present invention may be suitable as primers for PCR-based amplification techniques. Moreover, the oligonucleotides may be used for RNA interference (RNAi) approaches in order to modulate and, preferably down-regulate, the activity  
35

of the polypeptides encoded by the polynucleotides of the present invention. Thereby, an organism may be depleted of fatty acids and/or lipids and, specifically, a plant seed may be depleted of at least some of its seed storage compounds. As used herein, the term "RNA interference (RNAi)" refers to selective intracellular degradation of RNA used to silence expression of a selected target gene, i.e. the polynucleotide of the present invention. RNAi is a process of sequence-specific, post-transcriptional gene silencing in organisms initiated by double-stranded RNA (dsRNA) that is homologous in sequence to the gene to be silenced. The RNAi technique involves small interfering RNAs (siRNAs) that are complementary to target RNAs (encoding a gene of interest) and specifically destroy the known mRNA, thereby diminishing or abolishing gene expression. RNAi is generally used to silence expression of a gene of interest by targeting mRNA, however, any type of RNA is encompassed by the RNAi methods of the invention. Briefly, the process of RNAi in the cell is initiated by long double stranded RNAs (dsRNAs) being cleaved by a ribonuclease, thus producing siRNA duplexes. The siRNA binds to another intracellular enzyme complex which is thereby activated to target whatever mRNA molecules are homologous (or complementary) to the siRNA sequence. The function of the complex is to target the homologous mRNA molecule through base pairing interactions between one of the siRNA strands and the target mRNA. The mRNA is then cleaved approximately 12 nucleotides from the 3' terminus of the siRNA and degraded. In this manner, specific mRNAs can be targeted and degraded, thereby resulting in a loss of protein expression from the targeted mRNA. A complementary nucleotide sequence as used herein refers to the region on the RNA strand that is complementary to an RNA transcript of a portion of the target gene. The term "dsRNA" refers to RNA having a duplex structure comprising two complementary and anti-parallel nucleic acid strands. Not all nucleotides of a dsRNA necessarily exhibit complete Watson-Crick base pairs; the two RNA strands may be substantially complementary. The RNA strands forming the dsRNA may have the same or a different number of nucleotides, with the maximum number of base pairs being the number of nucleotides in the shortest strand of the dsRNA. Preferably, the dsRNA is no more than 49, more preferably less than 25, and most preferably between 19 and 23, nucleotides in length. dsRNAs of this length are particularly efficient in inhibiting the expression of the target gene using RNAi techniques. dsRNAs are subsequently degraded by a ribonuclease enzyme into short interfering RNAs (siRNAs). RNAi is mediated by small interfering RNAs (siRNAs). The term "small interfering RNA" or "siRNA" refers to a nucleic acid molecule which is a double stranded RNA agent that is complementary to i.e., able to base-pair with, a portion of a target RNA (generally mRNA), i.e. the polynucleotide of the present invention being RNA. siRNA acts to specifically guide en-

zymes in the host cell to cleave the target RNA. By virtue of the specificity of the siRNA sequence and its homology to the RNA target, siRNA is able to cause cleavage of the target RNA strand, thereby inactivating the target RNA molecule. Preferably, the siRNA which is sufficient to mediate RNAi comprises a nucleic acid sequence comprising an inverted repeat fragment of the target gene and the coding region of the gene of interest (or portion thereof). Also preferably, a nucleic acid sequence encoding a siRNA comprising a sequence sufficiently complementary to a target gene is operatively linked to an expression control sequence. Thus, the mediation of RNAi to inhibit expression of the target gene can be modulated by said expression control sequence. Preferred expression control sequences are those which can be regulated by an exogenous stimulus, such as the tet operator whose activity can be regulated by tetracycline or heat inducible promoters. Alternatively, an expression control sequence may be used which allows tissue-specific expression of the siRNA. The complementary regions of the siRNA allow sufficient hybridization of the siRNA to the target RNA and thus mediate RNAi. In mammalian cells, siRNAs are approximately 21-25 nucleotides in length (see Tuschl et al. 1999 and Elbashir et al. 2001). The siRNA sequence needs to be of sufficient length to bring the siRNA and target RNA together through complementary base-pairing interactions. The siRNA used with the Tet expression system of the invention may be of varying lengths. The length of the siRNA is preferably greater than or equal to ten nucleotides and of sufficient length to stably interact with the target RNA; specifically 15-30 nucleotides; more specifically any integer between 15 and 30 nucleotides, most preferably 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, and 30. By "sufficient length" is meant an oligonucleotide of greater than or equal to 15 nucleotides that is of a length great enough to provide the intended function under the expected condition. By "stably interact" is meant interaction of the small interfering RNA with target nucleic acid (e.g., by forming hydrogen bonds with complementary nucleotides in the target under physiological conditions). Generally, such complementarity is 100% between the siRNA and the RNA target, but can be less if desired, preferably 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%. For example, 19 bases out of 21 bases may be base-paired. In some instances, where selection between various allelic variants is desired, 100% complementary to the target gene is required in order to effectively discern the target sequence from the other allelic sequence. When selecting between allelic targets, choice of length is also an important factor because it is the other factor involved in the percent complementary and the ability to differentiate between allelic differences. Methods relating to the use of RNAi to silence genes in organisms, including *C. elegans*, *Drosophila*, plants, and mammals, are known in the art (see, for example, Fire et al., *Nature* (1998) 391:806-811; Fire, *Trends*

Genet. 15, 358-363 (1999); Sharp, RNA interference 2001. Genes Dev. 15,485-490 (2001); Hammond et al. Nature Rev. Genet. 2, 1110-1119 (2001); Tuschl, Chem. Biochem. 2, 239-245 (2001); Hamilton et al., Science 286, 950-952 (1999); Hammond et al., Nature 404, 293-296 (2000); Zamore et al., Cell 101, 25-33 (2000); Bernstein et al., Nature 409, 363-366 (2001); Elbashir et al., Genes Dev. 15, 188-200 (2001); WO 0129058; WO 09932619; and Elbashir et al., 2001 Nature 411: 494-498).

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.

Additionally, the present invention relates to and provides the use of the polynucleotides of the present invention 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 polynucleotide of the present invention, 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.

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 transgenic overexpression, co-suppression approaches, antisense approaches, and in vivo mutagenesis of the LMP nucleic acid or micro-RNA based techniques. 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.

More specifically, the present invention includes and provides a method for altering (increasing or decreasing or changing the specific profile) of the total oil content in a seeds

comprising: Transforming a plant with a nucleic acid construct that comprises as operably linked components, a promoter and nucleic acid sequences capable of modulating the level of the polynucleotides or polypeptides of the present invention, and growing the plant. Furthermore, the present invention includes and provides a method for altering (increasing or decreasing) the level of oleic acid in a seed comprising: transforming a plant with a nucleic acid construct that comprises as operably linked components, a promoter, a structural nucleic acid sequence capable of altering (increasing or decreasing) the level of oleic acid, and growing the plant

- 5
- 10 Also included herein is a seed produced by a transgenic plant transformed by the polynucleotides of the present invention, wherein the seed contains the said polynucleotide and wherein the plant is true breeding for a modified level of a seed storage compound. The present invention additionally includes seed oil produced by the aforementioned seed.
- 15 Further provided by the present invention are vectors comprising the polynucleotides of the present invention, host cells containing the vectors, and descendent plant materials produced by transforming a plant cell with the nucleic acids and/or vectors.

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.

It is further provided a method of producing a higher or lower than normal or typical level of storage compound in a transgenic plant expressing the polynucleotides of the present invention from *Arabidopsis thaliana* in the transgenic plant, wherein the transgenic plant is *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Oryza sativa*, *Zea mays*, *Triticum aestivum*, *Helianthus annuus* or *Beta vulgaris* or a species different from *Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Oryza sativa*, *Zea mays*, *Triticum aestivum*, *Helianthus annuus* or *Beta vulgaris*. 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

Arabidopsis thaliana, Brassica napus, Glycine max, Oryza sativa, Zea mays, Triticum aestivum, Helianthus annuus or Beta vulgaris is used, this also means Arabidopsis thaliana and/or Brassica napus and/or Glycine max and/or Oryza sativa and/or Triticum aestivum and/or Zea mays and/or Helianthus annuus and/or Beta vulgaris.

5

Accordingly, it is an object of the present invention to provide novel polynucleotides encoding LMPs as well as the corresponding polypeptides from Arabidopsis thaliana 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.

10

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.

15

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 may be useful for increasing plant yield under stress conditions (drought, light, cold, UV) by modulating light utilization efficiency. Moreover, polynucleotides and polypeptides of the present invention will improve seed germination and seed dormancy and, hence, will improve plant growth and/or yield of seed storage compounds.

20

25

The polynucleotides of the present invention may further comprise an operably linked promoter or partial promoter region. The promoter can be a constitutive promoter, an inducible promoter, or a tissue-specific promoter. The constitutive promoter can be, for example, the superpromoter (Ni et al., Plant J. 7:661-676, 1995; US5955646) or the PtxA promoter (WO 05/085450, Song H. et al.). The tissue-specific promoter can be active in vegetative tissue or reproductive tissue. The tissue-specific promoter active in reproductive tissue can be a seed-specific promoter. The tissue-specific promoter active in vegetative tissue can be a root-specific, shoot-specific, meristem-specific, or leaf-specific promoter. The polynucleo-

30

tides of the present invention can still further comprise a 5' non-translated sequence, 3' non-translated sequence, introns, or the combination thereof.

5 The present invention also provides a method for altering (increasing or decreasing) the number and/or size of one or more plant organs of a plant expressing a polynucleotide of the present invention, preferably, from *Arabidopsis thaliana* encoding a polypeptide of the present invention. More specifically, seed size and/or seed number and/or weight might be manipulated. Moreover, root length can be increased. Longer roots can alleviate not only the effects of water depletion from soil but also improve plant anchorage/standability, thus  
10 reducing lodging. Also, longer roots have the ability to cover a larger volume of soil and improve nutrient uptake. All of these advantages of altered root architecture have the potential to increase crop yield. Additionally, the number and size of leaves might be increased by the nucleic acid sequences provided in this application. This will have the advantage of improving photosynthetic light utilization efficiency by increasing photosynthetic  
15 light-capture capacity and photosynthetic efficiency.

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

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

Before the present compounds, compositions, methods and preferred embodiments thereof are disclosed and described in more detail, it is to be understood that this invention is not  
25 limited to specific polynucleotides, 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  
30 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 up to a plurality of cells can be utilized.

The present invention is based, in part, on the isolation and characterization of nucleic acid molecules a phosphate transporter LMP from plants including Arabidopsis, canola (*Brassica napus*) and other related crop species like maize, barley, linseed, sugar beet, or sunflower.

- 5 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 plant (*Brassica napus*) encoding a Lipid Metabolism Protein (LMP), or a portion thereof.

One aspect of the invention pertains to isolated nucleic acid molecules that encode LMP polypeptides or biologically active portions thereof, as well as 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.

10 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 which 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 that 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

15 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 *Brassica napus* 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

20 chemical precursors or other chemicals when chemically synthesized. A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having a nucleotide sequence of the polynucleotide of the present invention, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. For example, a *Arabidopsis thaliana* or *Brassica napus* LMP cDNA can be isolated from an a

25 *Arabidopsis thaliana* or *Brassica napus* library using all or portion of one of the sequences

30

35

of the polynucleotide of the present invention 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 SEQ ID NO:1 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 SEQ ID NO:1 can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this same sequence of SEQ ID NO: 1). 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 SEQ ID NO: 1. 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.

In a preferred embodiment, an isolated nucleic acid of the invention comprises one of the nucleotide sequences shown of the polynucleotide of the present invention. The sequence of SEQ ID NO: 1 corresponds to the *Arabidopsis thaliana* LMP cDNA of the invention. These cDNAs comprise sequences encoding LMPs (i.e., the "coding region", indicated in SEQ ID NO: 1), 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 SEQ ID NO: 1 or can contain whole genomic fragments isolated from genomic DNA.

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 SEQ ID NO: 1, or a portion thereof. A nucleic acid molecule which is complementary to one of the nucleotide sequences shown in SEQ ID NO: 1 is one which is sufficiently

complementary to one of the nucleotide sequences shown in SEQ ID NO: 1 such that it can hybridize to one of the nucleotide sequences shown in SEQ ID NO: 1, thereby forming a stable duplex. 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 SEQ ID NO: 1, or a portion thereof. Specific algorithms for the determination of the degree of identity are found elsewhere in this specification. 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 SEQ ID NO: 1, 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. Specific hybridization conditions are to be found elsewhere in this specification. Moreover, the nucleic acid molecule of the invention can comprise only a portion of the coding region of one of the sequences in SEQ ID NO: 1, 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 gene from *Arabidopsis thaliana* 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 SEQ ID NO: 1, an anti-sense sequence of one of the sequences set forth in SEQ ID NO: 1, or naturally occurring mutants thereof. Primers based on a nucleotide sequence of SEQ ID NO: 1 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. 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 SEQ ID NO: 2 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 SEQ ID NO: 2) 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. How to determine the degree of identical or equivalent amino acids between two sequences is set forth elsewhere in this specification in detail. Transporter proteins, such as the phosphate transporter, play a role in the biosynthesis of seed storage compounds. Examples of such activities are described herein. Examples of LMP-encoding nucleic acid sequences are set forth in SEQ ID NO: 1 and SEQ ID NO: 5.

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, sugar beet 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.

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 has an activity as set forth above. 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.

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 SEQ ID NO: 1 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 and in accordance with the present invention, preferably, the phosphate transporter activity. 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. 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.

The invention further encompasses nucleic acid molecules that differ from one of the nucleotide sequences shown in SEQ ID NO: 1 (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 SEQ ID NO: 1. 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 SEQ ID NO: 1. In one embodiment, the full-length nucleic acid or protein or fragment of the nucleic acid or protein is from *Arabidopsis thaliana*. In addition to the LMP nucleotide sequences shown in SEQ ID NO: 1, 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 *Arabidopsis thaliana* 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 *Arabidopsis thaliana* 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. Nucleic acid molecules corresponding to natural variants and non-*Arabidopsis thaliana* orthologs of the LMP cDNA of the invention can be isolated based on their homology LMP nucleic acid disclosed herein

5 using the 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

10 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 SEQ ID NO: 1. 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

15 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.,

20 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 SEQ ID NO: 1 or 3 corresponds to a naturally occurring nucleic acid molecule. As used herein, a

25 "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 *Arabidopsis thaliana* LMP. 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 SEQ ID NO: 1, thereby leading to changes in the amino acid sequence of the

30 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 SEQ ID NO: 2. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of one of the LMPs (SEQ ID NO: 2) without

35 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.

5 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  
10 by a nucleic acid of SEQ ID NO: 1 and has one or more activities set forth above. 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 SEQ ID NO: 1, more preferably at least about 60-70% homologous to one of the sequences encoded by a nucleic acid of SEQ  
15 ID NO: 1, even more preferably at least about 70-80%, 80-90%, 90-95% homologous to one of the sequences encoded by a nucleic acid of SEQ ID NO: 1, and most preferably at least about 96%, 97%, 98%, or 99% homologous to one of the sequences encoded by a nucleic acid of SEQ ID NO: 1.

20 To determine the percent homology of two amino acid sequences (e.g., one of the sequences encoded by a nucleic acid of SEQ ID NO: 1 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  
25 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 SEQ ID NO: 1) 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 SEQ ID NO: 1), then the molecules are homologous at that position (i.e., as used herein amino acid or nucleic acid "homology" is equivalent to amino acid  
30 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).

An isolated nucleic acid molecule encoding a LMP homologous to a protein sequence encoded by a nucleic acid of SEQ ID NO: 1 can be created by introducing one or more nu-  
35

cleotide substitutions, additions or deletions into a nucleotide sequence of SEQ ID NO: 1 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 SEQ ID NO: 1 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 SEQ ID NO: 1, the encoded protein can be expressed recombinantly and the activity of the protein can be determined using, for example, assays described herein (see Examples 11-13 of the Exemplification).

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.

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 se-

quence 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  
5 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-  
10 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.

15  
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,  
20 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 that give rise to complementary overhangs between  
25 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  
30 fusion moiety is linked in-frame to the LMP.

In addition to the nucleic acid molecules encoding LMPs described above, another aspect of the invention pertains to isolated nucleic acid molecules that are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a  
35 "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a dou-

ble-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can be 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 that are translated into amino acid residues. 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 that flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding LMP disclosed herein (e.g., the sequences set forth in SEQ ID NO:1), 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 that 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).

10 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  
15 to separate the sense and antisense fragments of LMP sequences in the construct.

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  
20 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.  
25  
30

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  
35 run parallel to each other (Gaultier et al. 1987, Nucleic Acids Res. 15:6625-6641). The an-

tisense 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).

5 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 a  
10 LMP-encoding nucleic acid can be designed based upon the nucleotide sequence of a LMP cDNA disclosed herein 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 having a specific ribonuclease activity from a pool of RNA molecules (see, e.g., Bartel, D. & Szostak J.W. 1993, *Science* 261:1411-1418).

20 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).

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"  
30 refers to a nucleic acid molecule capable of transporting 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  
35 (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vec-

tors). 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.

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 that direct constitutive expression of a nucleotide sequence in many types of host cell and those that 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.).

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, Glaucoma, Platyophrya, Potomacus, Pseudocohnilembus, Euplotes, Engelmaniella, 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.

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 purifi-

cation of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin, and enterokinase.

Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith & Johnson  
5 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  
10 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.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et  
15 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 po-  
20 lymerase (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.

One strategy to maximize recombinant protein expression is to express the protein in host  
25 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  
30 (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.

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,  
35 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.

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).

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, Fritsch and Maniatis, *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

In another embodiment, the LMPs of the invention may be expressed in uni-cellular 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 and R. Wu, Academic Press, 1993, S. 15-38).

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.

- 5 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).

10

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  
15 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.  
20 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  
25 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  
30 *Sorghum kasirin*-gene, and the rye secalin gene).

Plant gene expression can also be facilitated via an inducible promoter (for a 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.  
35 Examples for such promoters are a salicylic acid inducible promoter (WO 95/19443), a tet-

racycline inducible promoter (Gatz et al. 1992, *Plant J.* 2:397-404), and an ethanol inducible promoter (WO 93/21334).

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).

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.

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 that 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).

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.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "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.

For stable transfection of mammalian and plant cells, it is known that, depending upon the used expression vector and transfection technique, 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).

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 *Arabidopsis thaliana* or *Brassica napus* 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 chimeroplasty (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.

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.

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 ex-

pression of the LMP gene only in the presence of IPTG. Such regulatory systems are well known in the art.

5 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  
10 method further comprises isolating LMPs from the medium or the host cell.

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  
15 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  
20 "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%  
25 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 that 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  
30 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  
35 LMP is derived. Typically, such proteins are produced by recombinant expression of, for

example, an *Arabidopsis thaliana* or *Brassica napus* LMP in other plants than *Arabidopsis thaliana* or *Brassica napus* or microorganisms, algae or fungi.

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 *Brassica napus* or of cellular membranes, or has one or more of the activities set forth above. 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 SEQ ID NO: 1 such that the protein or portion thereof maintains its phosphate transporter activity.

10 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 SEQ ID NO: 1. 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 SEQ ID NO: 1. 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 SEQ ID NO: 1. The preferred LMPs of the present invention also

20 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 SEQ ID NO: 1, and which has one or more of the activities set forth above.

25 In other embodiments, the LMP is substantially homologous to an amino acid sequence encoded by a nucleic acid of SEQ ID NO: 1 and retains the functional activity of the protein of one of the sequences encoded by a nucleic acid of SEQ ID NO: 1 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 full *Arabidopsis thaliana* protein which is substantially homologous to an entire amino acid sequence encoded by a nucleic acid of SEQ ID NO: 1.

35

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).

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 that 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.

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 that 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 degen-

erate 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).

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.

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 that 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).

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

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 *Arabidopsis thaliana* and related organisms; mapping of genomes of organisms related to *Arabidopsis thaliana*; identification and localization of *Arabidopsis thaliana* 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.

10 Higher plants like *Arabidopsis thaliana* or *Brassica napus* 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  
15 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.

The LMP nucleic acid molecules of the invention have a variety of uses. First, the nucleic acid and protein molecules of the invention may serve as markers for specific regions of the  
20 genome. This has utility not only in the mapping of the genome, but also for functional studies of *Arabidopsis thaliana* or *Brassica napus* proteins. For example, to identify the region of the genome, to which a particular *Arabidopsis thaliana* or *Brassica napus* DNA-binding protein binds, the *Arabidopsis thaliana* or *Brassica napus* genome could be digested, and  
25 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 *Arabidopsis thaliana* or *Brassica napus*, and, when performed multiple times with different enzymes, facilitates a rapid  
30 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.

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.

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.

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 overexpression 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 signal-

ing 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'.

5 The present invention also provides antibodies that specifically bind to an LMP-polypeptide, or a portion thereof, as encoded by a nucleic acid disclosed herein or as described herein.

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  
10 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.  
15 1992, *Bio/Technology* 10:169-175).

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  
20 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 selec-  
25 tively 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.

In some instances, it is desirable to prepare monoclonal antibodies from various hosts. A  
30 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).

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.

5

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.

10

## EXAMPLES

### Example 1 - General Processes

15

a) General Cloning Processes. 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).

20

b) Chemicals. 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.

30

35

c) Plant Material and Growth: Arabidopsis plants. For this study, root material, leaves, siliques and seeds of wild-type and mutant plants of Arabidopsis thaliana were used. The ctr1 mutant was isolated from Columbia ecotype as described (Kieber JJ et al., Cell 72:427-441). Wild type and ctr1 Arabidopsis 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). All plants were started on half-strength MS medium (Murashige & Skoog, 1962, Physiol. Plant. 15, 473-497), pH 6.2, 2% sucrose and 1.2% agar. Seeds were sterilized for 20 minutes in 20% bleach 0.5% triton X100 and rinsed 6 times with excess sterile water. Plants were either grown as described above or on soil under standard conditions as described in Focks & Benning (1998, Plant Physiol. 118:91-101).

#### Example 2 - Total DNA Isolation from Plants.

The details for the isolation of total DNA relate to the working up of 1 gram fresh weight of plant material.

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.

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  $\mu\text{l}$  of N-laurylsarcosine buffer, 20  $\mu\text{l}$  of  $\beta$ -mercaptoethanol and 10  $\mu\text{l}$  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 10,000 g for 30 min and resuspended in 180  $\mu\text{l}$  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  $\mu\text{l}$  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*.

5

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

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

RNA preparation from *Arabidopsis* seeds - "hot" extraction:

10

#### 1. Buffers, enzymes and solution

2M KCl

Proteinase K

15

Phenol (for RNA)

Chloroform:Isoamylalcohol

(Phenol:chloroform 1:1; pH adjusted for RNA)

4 M LiCl, DEPC-treated

DEPC-treated water

20

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

25

Extraction Buffer:

0.2M Na Borate

30 mM EDTA

30 mM EGTA

1% SDS (250µl of 10% SDS-solution for 2.5ml buffer)

30

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 β-Mercaptoethanol (fresh, handle under fume hood - use 35µl of 14.3M so-

35

lution for 5ml buffer)

2. Extraction. 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.

5 Add 350µl preheated extraction buffer (here for 100mg tissue, buffer volume can be as much as 500µl 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.

10 3. Digestion. Add Proteinase K (0.15mg/100mg tissue), vortex and keep at 37°C for one hour.

First Purification. 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 15 min at RT). Resolve pellet in 1ml 4M LiCl by 10 to 15min vortexing. Pellet RNA by 5min centrifugation.

20 Second Purification. 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 chloroform: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 25 1:200 dilution at 260 and 280nm. 40µg RNA/ml = 1OD260

30 RNA from wild-type and the wri1 mutant of 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).

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

35 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)<sup>+</sup> 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.

Total RNA was extracted from tissues using RNeasy Maxi kit (Qiagen) according to manufacturer's protocol and mRNA was processed from total RNA using Oligotex mRNA Purification System kit (Qiagen), also according to manufacturer'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.

10

#### Example 4 - cDNA Library Construction.

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.

Brassica napus 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

35

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 - Cloning of full-length cDNAs and orthologs of identified LMP genes.

5

Clones corresponding to full-length sequences and partial cDNAs from *Arabidopsis thaliana* had been identified in the in-house proprietary Hyseq databases. The Hyseq clones of *Arabidopsis thaliana* were sequenced at DNA Landmarks using a ABI 377 slab gel sequencer and BigDye Terminator Ready Reaction kits (PE Biosystems, Foster City, CA).  
10 Sequence alignments were done to determine whether the Hyseq clones were full-length or partial clones. In cases where the Hyseq clones were determined to be partial cDNAs the following procedure was used to isolate the full-length sequences. Full-length cDNAs were 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 RACE PCR primers  
15 were designed based on the Hyseq clone sequences. The isolation of full-length cDNAs and the RACE PCR protocol used were based on the manufacturer's conditions. The RACE product fragments were 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 were transformed into TOP10 cells (Invitrogen) using  
20 standard conditions (Sambrook et al. 1989). Transformed cells were 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 were 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)  
25 following manufacturer's instructions. Subsequent analyses of clones, and restriction mapping, was performed according to standard molecular biology techniques (Sambrook et al. 1989).

Full-length cDNAs were isolated and cloned into binary vectors by using the following procedure: Gene specific primers were designed using the full-length sequences obtained from  
30 Hyseq clones or subsequent RACE amplification products. Full-length sequences and genes were amplified utilizing Hyseq clones or cDNA libraries as DNA template 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  
35 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, 6°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 5 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 of *Arabidopsis*. 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 Mini-Prep DNA preparation kit. The insert was verified throughout the various cloning steps by 10 determining its size through restriction digest and inserts were sequenced to ensure the expected gene was used in *Arabidopsis* transformation.

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 15 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.

Gene sequences can be used to identify homologues or orthologs from cDNA or genomic 20 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 25 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.

30 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.

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' end of two complementary oligonucleotides with T4 polynucleotide kinase.

5 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.

10 Oligonucleotide hybridization solution:

6 x SSC

0.01M sodium phosphate

1 mM EDTA (pH 8)

15 0.5 % SDS

100 µg/ml denaturated salmon sperm DNA

0.1 % nonfat dried milk

During hybridization, temperature is lowered stepwise to 5-10°C below the estimated oligonucleotide  $T_m$  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).

25

Example 6 - Identification of Genes of Interest by Screening Expression Libraries with Antibodies.

c-DNA clones can be used to produce recombinant protein for example in *E. coli* (e. g. 30 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 35 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).

#### 5 Example 7 - Northern-Hybridization.

For RNA hybridization, 20 µg of total RNA or 1 µg of poly-(A)+ 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+, 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.

#### 20 Example 8 - DNA Sequencing and Computational Functional Analysis.

cDNA libraries can be 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 can be 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 can be 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 Qiagene DNA preparation robot (Qiagen, Hilden) according to the manufacturer's protocols). Sequences can be 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>.

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).

#### 25 Example 9 - Plasmids for Plant Transformation.

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' to the cDNA a plant promoter activates transcription of the cDNA. A polyadenylation sequence is located 3' 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' 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. 35 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.

5 Further examples for plant binary vectors are the pBPS-GB1, pSUN2-GW or pBPS-GB047 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 and a USP seed-specific promoter or the PtxA 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  
10 seed-specific or PtxA promoters. 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  
15 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).

20 Example 10 - Agrobacterium Mediated Plant Transformation.

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  
25 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.

30 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  
35 8:238-242; De Block et al. 1989, Plant Physiol. 91:694-701). Use of antibiotic for Agrobac-

terium 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 5 13:282-285).

The *Arabidopsis thaliana* phosphate transporter gene was cloned into a binary vector and expressed either under the USP promoter or the PtxA promoter (the promoter of the *Pisum sativum* PtxA gene), which is a promoter active in virtually all plant tissues. However, in 10 seeds and flowers, there is no expression activity detectable by GUS staining and low expression activity detectable with the more sensitive method of RT-PCR (Song, H-S. et al., WO 05/085450). Only in plant lines comprising multiple copies of a transgenic ptxA-promoter/GUS expression construct some expression could be detected in part of the flowers and the siliques (for more details see Song, H-S. et al., WO 05/085450). Alternatively, 15 the superpromoter, which is a constitutive promoter (Stanton B. Gelvin, US 5,428,147 and US 5,217,903) or seed-specific promoters like USP (unknown seed protein) from *Vicia faba* (Baeumlein et al. 1991, Mol. Gen. Genetics 225:459-67), 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. were used.

20 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 25 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 30 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.

35 The method of plant transformation is also applicable to other crops. In particular, seeds of canola are surface sterilized with 70% ethanol for 4 minutes at room temperature with con-

tinuous 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.

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). 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.

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 electro-

phoresed 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.

5 In general, a rice (or other monocot) phosphate transporter gene under a plant promoter like PtxA could be transformed into corn, or another crop plant, to generate effects of monocot phosphate transporter genes in other monocots, or dicot phosphate transporter genes in other dicots, or monocot genes in dicots, or vice versa. The plasmids containing these coding sequences, 5' of a promoter and 3' of a terminator would be constructed in a  
10 manner similar to those described for construction of other plasmids herein.

#### Example 11 - In vivo Mutagenesis.

In vivo mutagenesis of microorganisms can be performed by incorporation and passage of  
15 the plasmid (or other vector) DNA through *E. coli* or other microorganisms (e.g. *Bacillus* spp. or yeasts such as *Saccharomyces cerevisiae*) that 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-  
20 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.

25

#### Example 12 - Assessment of the mRNA Expression and Activity of a Recombinant Gene Product in the Transformed Organism.

The activity of a recombinant gene product in the transformed host organism can be meas-  
30 ured 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  
35 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).

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.

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.

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 13 - In vitro Analysis of the Function of Arabidopsis thaliana phosphate transporter Genes in Transgenic Plants.

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 struc-

ture, 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 14 - Purification of the Desired Product from Transformed Organisms.

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.

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.

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).

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), *Ullmann'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).

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: Ullmann'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).

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 Chromatogra-*

phy 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.

- 5 Unequivocal proof of the presence of fatty acid products (Table 1 and 2) 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  
 10 seeds by Focks & Benning (1998, Plant Physiol. 118:91-101).

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

15 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

The marked up (\*) fatty acids do not normally occur in plant seed oils, but their production in transgenic plant seed oil is of importance in plant biotechnology.

- 5 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).
- 10 Total seed oil levels can be measured by any appropriate method. Quantitation of seed oil contents is often performed with conventional methods, such as near infrared analysis (NIR) or nuclear magnetic resonance imaging (NMR). NIR spectroscopy has become a standard method for screening seed samples whenever the samples of interest have been amenable to this technique. Samples studied include canola, soybean, maize, wheat, rice,
- 15 and others. NIR analysis of single seeds can be used (see e.g. Velasco et al., "Estimation of seed weight, oil content and fatty acid composition in intact single seeds of rapeseed" (*Brassica napus* L.) by near-infrared reflectance spectroscopy, "Euphytica," Vol. 106, 1999, pp. 79-85). NMR has also been used to analyze oil content in seeds (see e.g. Robertson & Morrison, "Analysis of oil content of sunflower seed by wide-line NMR," *Journal of the*
- 20 *American Oil Chemists Society*, 1979, Vol. 56, 1979, pp. 961-964, which is herein incorporated by reference in its entirety).

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  $^{14}\text{C}$ -acetate or  $^{14}\text{C}$ -pyruvate

25 (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).

5

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  $100^{\circ}\text{C}$ , cooled on ice and centrifuged again followed by extraction in 0.5 M sulfuric acid in methanol containing 2% dimethoxypropane for 1 hour at  $90^{\circ}\text{C}$  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  $170^{\circ}\text{C}$  and  $240^{\circ}\text{C}$  for 20 minutes and 5 min. at  $240^{\circ}\text{C}$ . The identity of resulting fatty acid methylesters is defined by the use of standards available from commercial sources (i.e., Sigma).

15

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-Derivatoren (Christie, Oily Press, Dundee, 1998).

20

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).

25

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  $70^{\circ}\text{C}$  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

35

in 200  $\mu$ 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  $\mu$ l of 1 N acetic acid and centrifugation for 5 min at 16,000 g, the supernatant is used for starch quantification.

- 5 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  
10 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.

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  
15 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  
20 2 h at 25°C, the homogenate is centrifuged at 16,000 g for 5 min and 200 ml 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.

- 25 Enzymatic assays of hexokinase and fructokinase are performed spectrophotometrically according to Renz et al. (1993, *Planta* 190:156-165), of phosphoglucosomerase, ATP-dependent 6-phosphofructokinase, pyrophosphate-dependent 6-phosphofructokinase, Fructose-1,6-bisphosphate aldolase, triose phosphate isomerase, glycerol-3-P dehydrogenase, phosphoglycerate kinase, phosphoglycerate mutase, enolase and pyruvate kinase  
30 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).

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).

5 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).

10 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.

15 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.

20 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  
25 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).

30 Example 15 - Analysis of the Impact of Recombinant Proteins on the Production of a Desired Seed Storage Compound.

Seeds from transformed *Arabidopsis thaliana* plants were analysed by gas chromatography (GC) for total oil content and fatty acid profile. *Arabidopsis* (ecotype Columbia-2) was used

to investigate the influence of ph001 (Phosphate transporter) (SEQ ID NO. 1) on seed storage compound accumulation, see also for more information Table 3.

Table 3. A table of the functions of the LMP

5

SEQ ID NOs	Sequence name	Species	Function	ORF position
1	Ph001	Arabidopsis thaliana	Phosphate transporter	1 to 2355

Total fatty acid content of seeds of control and transgenic plants were measured with bulked seeds (usually 5 mg seed weight) of a single plant. Two different types of controls have been used: C-24 (Columbia-24, an Arabidopsis ecotype found to accumulate high amounts of total fatty acids in seeds) and BPS empty (without LMP gene of interest) binary vector construct. The controls indicated in the tables below have been grown side by side with the transgenic lines. Differences in the total values of the controls are explained either by differences in the growth conditions, which were found to be very sensitive to small variations in the plant cultivation. Because of the seed bulking all values obtained with T3 seeds are either the result of a mixture of homozygous (for the gene of interest) and heterozygous events or are homozygous (zygosity has not been determined), implying that these data may potentially underestimate the LMP gene effect. Ph001 expression was driven by the PtxA promoter.

20 Table 4. Determination of the T3 seed total fatty acid content of transgenic lines of ph001 (containing SEQ ID NO: 1). Shown are the means ( $\pm$  standard deviation). Average mean values are shown  $\pm$  standard deviation, number of individual measurements per plant line: 6-8 for controls, 15-20 for transgenic lines; Col-2 is the Arabidopsis ecotype the LMP gene has been transformed in, C-24 is a high-oil Arabidopsis ecotype used as another control.

25 Transgenic seeds of ph001-26 and ph001-27 show a significant increase relative to the empty vector control seeds ( $p < 0.05$  as obtained by simple t-test).

Genotype	g total fatty acids/g seed weight
C-24 wild-type seeds	0.310 ± 0.008
Empty vector control seeds	0.266 ± 0.023
5 ph001-26 transgenic seeds	0.294 ± 0.026
ph001-27 transgenic seeds	0.301 ± 0.021
ph001-14 transgenic seeds	0.281 ± 0.022
ph001-21 transgenic seeds	0.272 ± 0.032

10 Table 5: Orthologs of Phosphate Transporter

Organism	SEQ ID NO:	Nucleotide Length bp	Nucleotide Identity ph001 %	Protein Length aa	Protein Identity ph001 %	Protein Similarity ph001 %
Arabidopsis thaliana	1	2355	100	784	100	100
Oryza sativa	5	5793	66	1930	61	75

15 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.

## CLAIMS

1. A polynucleotide comprising a nucleic acid sequence selected from the group consisting of:  
5  
  - (a) a nucleic acid sequence as shown in SEQ ID NO: 1;
  - (b) a nucleic acid sequence encoding a polypeptide having an amino acid sequence as shown in SEQ ID NO: 2;  
10
  - (c) a nucleic acid sequence which is at least 66% identical to the nucleic acid sequence of (a) or (b), wherein said nucleic acid sequence encodes a polypeptide having phosphate transporter activity; and
  - 15 (d) a nucleic acid sequence being a fragment of any one of (a) to (c), wherein said fragment encodes a polypeptide or biologically active portion thereof having phosphate transporter activity.
2. The polynucleotide of claim 1, wherein said polynucleotide is DNA or RNA.  
20
3. A vector comprising the polynucleotide of claim 1 or 2.
4. The vector of claim 3, wherein said vector is an expression vector.
- 25 5. A host cell comprising the polynucleotide of claim 1 or 2 in a vector of claim 3 or 4.
6. A method for the manufacture of a polypeptide having phosphate transporter activity comprising:  
  - (a) expressing the polynucleotide of claim 1 or 2 in a host cell; and  
30
  - (b) obtaining the polypeptide encoded by said polynucleotide from the host cell.
7. A polypeptide encoded by the polynucleotide of claim 1 or 2 or which is obtainable by the method of claim 6.
- 35 8. A transgenic non-human organism comprising the polynucleotide of claim 1 or 2, the vector of claim 3 or 4 or the host cell of claim 5.
9. The transgenic non-human organism of claim 9, wherein said non-human transgenic organism is a plant.  
40
10. A transgenic seed produced by a plant according to claim 9.

11. Use of a transgenic plant according to claim 9 or a transgenic seed according to claim 10 for the manufacture of an agricultural product.
- 5 12. A method for the manufacture of a lipid or a fatty acid comprising the steps of:
- (a) cultivating the host cell of claim 5 or the transgenic non-human organism of claim 8 or 9 under conditions allowing synthesis of the said lipid or fatty acid; an
  - (b) obtaining the said lipid or fatty acid from the host cell or the transgenic non-human organism.
- 10
13. A method for the manufacture of a plant having a modified amount of a seed storage compound comprising the steps of:
- (a) introducing the polynucleotide of claim 1 or 2 or the vector of claim 3 or 4 into a plant cell; and
  - 15 (b) generating a transgenic plant from the said plant cell, wherein the polypeptide encoded by the polynucleotide modifies the amount of the said seed storage compound in the transgenic plant.
- 20
14. The method of claim 13, wherein the amount of said seed storage compound is increased compared to a non-transgenic control plant.
15. The method of claim 13 or 14, wherein said seed storage compound is a lipid or a fatty acid.