(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 February 2007 (01.02.2007)

PCT

(10) International Publication Number
WO 2007/012576 A2

(51) International Patent Classification:
C12N 15/82 (2006.01) *A01H 5/00* (2006.01)

(21) International Application Number:
PCT/EP2006/064276

(22) International Filing Date: 14 July 2006 (14.07.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/595,649 25 July 2005 (25.07.2005) US

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

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **ZANK, Thorsten** [DE/DE]; Seckenheimer Str. 4-6, 68165 Mannheim (DE). **OSWALD, Oliver** [DE/DE]; Nibelungenstr. 158, 64686 Lautertal (DE).

(74) Agent: **KOCK, Michael**; BASF Aktiengesellschaft, 67056 Ludwigshafen (DE).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, 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, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) **Title:** COMBINATION OF LIPID METABOLISM PROTEINS AND USES THEREOF

Open reading frames of the nucleic acid sequences coding for LMP usefull in novel combinations to increase oil content:

1A: Wri (SEQ ID NO. 1):

atgaagaagcgcttaaccacttccacttgttcttcttctccatcttcctctgtttcttcttctactactacttcctctcctattcagtcggagg
ctccaaggcctaaacgagccaaaagggctaagaaatcttctccttctggtgataaatctcataacccgacaagccctgcttcta
cccgacgcagctctatctacagaggagtcactagacatagatggactgggagattcgaggctcatctttgggacaaaagctctt
ggaattcgattcagaacaagaaaggcaaacaagtttatctgggagcatatgacagtgaagaagcagcagcacatacgtacg
atctggctgctctcaagtactggggacccgacaccatcttgaattttccggcagagacgtacacaaaggaattggaagaaatg
cagagagtgacaaaggaagaatatttggcttctctccgccgccagagcagtggtttctccagaggcgtctctaaatatcgcggc
gtcgctaggcatcaccacaacggaagatgggaggctcggatcggaagagtgtttgggaacaagtactgtacctcggcacct
ataatacgcaggaggaagctgctgcagcatatgacatggctgcgattgagtatcgaggcgcaaacgcggttactaatttcgac
attagtaattacattgaccggttaaagaagaaaggtgtttttcccgttccctgtgaaccaagctaaccatcaagagggtattcttgtt
gaagccaaacaagaagttgaaacgagagaagcgaaggaagagcctagagaagaagtgaaacaacagtacgtggaag
aaccaccgcaagaagaagaagagaaggaagaagagaaagcagagcaacaagaagcagagattgtaggatattcaga
agaagcagcagtggtcaattgctgcatagactcttcaaccataatggaaatggatcgttgtggggacaacaatgagctggctg
gaacttctgtatgatggatacagggttttctccgttttttgactgatcagaatctcgcgaatgagaatcccatagagtatccggagcta
ttcaatgagttagcatttgaggacaacatcgacttcatgttcgatgatgggaagcacgagtgcttgaacttggaaaatctggattgt
tgcgtggtgggaagagagagcccaccctcttcttcttccaccattgtcttgcttatctactgactctgcttcatcaacaacaacaaca
acaacctcggtttcttgtaactatttggtctga

1B JB05 (SEQ ID NO. 2):

atggacggtgccggagaatcacgactcggtggtgatggtggtggtgatggttctgttggagttcagatccgacaaacacggatg
ctaccggatttttctccagagcgtgaatctcaagtatgtgaaattaggttaccattacttaatctcaaaictcttgactctctgtttattccc
tctcgccgttgttatctccgtcgaagcctctcagatgaacccagatgatctcaaacagctctggatccatctacaatacaatctggt
tagtatcatcatctgttcagcgattctagtcttcgggttaacggtttatgttatgaccccgacctagacccgtttacttggttgatttctcttg
ttatctcccacctgatcacctcaaagctcctacgctcggttcatggaacatttctagactcaccggagatttcgatgactctgctctc
gagtttcaacgcaagatccttgagcgttctggtttaggggaagacacttatgtcccgaagctatgcattatgttccaccgagaatttt
caatggctgctgctagagaagaagctgaacaagtcatgtttggtgctttagataaccttttcgctaacactaatgtgaaaccaaa
ggatattggaatccttgttgtgaattgtagtctcttaatccaactccttcgttatctgcaatgattgtgaacaagtataagcttagaggt
aacattagaagctacaatctaggcggtatgggttgcagcgcgggagttatcgctgtggatcttgctaaagacatgttgttggtaca
taggaacacttatgcggttgttgtttctactgagaacattactcagaattggtatttttggtaacaagaaatcgatgttgatacccgaact
gcttgtttcgagttggtggctctgcggttttgctatcgaacaagtcgagggacaagagacggtctaagtacaggcttgtacatgta
gtcaggactcaccgtggagcagatgataaagcttttccgttgtgtttatcaagagcaggatgatacagggagaaccgggtttccg
ttgtcgaaagatctaatggcgattgcaggggaaactctcaaaaccaatatcactacattgggtcctcttgttctaccgataagtga
gcagattcccttctttatgactctagttgtgaagaagctctttaacggtaaagtgaaaccgtatatcccggatttcaaacttgctttcg
agcatttctgtatccatgctggtggaagagctgtgatcgatgagttagagaagaaatctgcagcttttcaccagttcatgtcgaggctt
cgaggatgactcttcatcgatttggtaacacatcttcgagctccatttggtatgaattggcttacattgaagcgaagggaaggatg
cgaagaggtaatcgtgtttggcaaatcgcgttcggaagtggatttaaatgtaatagcgcgatttgggaagcattaaggcatgtga
aaccttcgaacaacagtccttgggaagattgtattgacaagtatccggtaacttaagttattag

1C JB4054 (SEQ ID NO. 3):

atgtgttcattagagaaacgtgatcgtctttttcatactaaaactcaccggcgacggcgaacaccgtctaaacccaaccttattcga
ctctctccgctccaccatcaaccaaatccgatcagatccatcatttttcacaatcagtactcatcacaacatcagatggtaaattctt
ctccaacggctacgatctcgctttagccgagtcaaatccttctctctctgttgtaatggacgcaaaacttagatccttagtcgccgat
ctaatctctcttcctatgccaacaatcgccgccgtcacaggtcacgcttccgccgcgggatgtattttagcgatgagtcatgattat

(57) **Abstract:** Described herein are inventions in the field of genetic engineering of plants, including combinations of nucleic acid molecules encoding LMPs to improve agronomic, horticultural, and quality traits. This invention relates generally to the combination of nucleic acid sequences encoding proteins that are related to the presence of seed storage compounds in plants. More specifically, the present invention relates to LMP nucleic acid sequences encoding lipid metabolism proteins (LMP) and the use of these combinations of these sequences, their order and direction in the combination, and the regulatory elements used to control expression and transcript termination in these combinations in transgenic plants. In particular, the invention is directed to methods for manipulating fatty acid-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 combinations of polypeptides to stimulate plant growth and/or to increase yield and/or composition of seed storage compounds.

WO 2007/012576 A2

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*

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

Combination of Lipid Metabolism Proteins and Uses Thereof

Field of the Invention

Described herein are inventions in the field of genetic engineering of plants, including combinations of nucleic acid molecules encoding LMPs to improve agronomic, horticultural, and quality traits. This invention relates generally to the combination of nucleic acid sequences encoding proteins that are related to the presence of seed storage compounds in plants. More specifically, the present invention relates to LMP nucleic acid sequences encoding lipid metabolism proteins (LMP) and the use of these combinations of these sequences, their order and direction in the combination, and the regulatory elements used to control expression and transcript termination in these combinations in transgenic plants. In particular, the invention is directed to methods for manipulating fatty acid-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 combinations of polypeptides to stimulate plant growth and/or to increase yield and/or composition of seed storage compounds.

Background of the Invention

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

Plant seed oils comprise both neutral and polar lipids (see Table 2). The neutral lipids contain primarily triacylglycerol, which is the main storage lipid that accumulates in oil bodies in seeds. The polar lipids are mainly found in the various membranes of the seed cells, e.g. the endoplasmic reticulum, microsomal membranes, plastidial and mitochondrial membranes and the cell membrane. The neutral and polar lipids contain several common fatty acids (see Table

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

Lipids are synthesized from fatty acids and their synthesis may be divided into two parts: the prokaryotic pathway and the eukaryotic pathway (Browse et al. 1986, Biochemical J. 235:25-31; Ohlrogge & Browse 1995, Plant Cell 7:957-970). The prokaryotic pathway is located in plastids that are also the primary site of fatty acid biosynthesis. Fatty acid synthesis begins with the conversion of acetyl-CoA to malonyl-CoA by acetyl-CoA carboxylase (ACCase). Malonyl-CoA is converted to malonyl-acyl carrier protein (ACP) by the malonyl-CoA:ACP transacylase. The enzyme beta-keto-acyl-ACP-synthase III (KAS III) catalyzes a condensation reaction, in which the acyl group from acetyl-CoA is transferred to malonyl-ACP to form 3-ketobutyryl-ACP. In a subsequent series of condensation, reduction and dehydration reactions the nascent fatty acid chain on the ACP cofactor is elongated by the step-by-step addition (condensation) of two carbon atoms donated by malonyl-ACP until a 16- or 18-carbon saturated fatty acid chain is formed. The plastidial delta-9 acyl-ACP desaturase introduces the first double bond into the fatty acid.

In the prokaryotic pathway the saturated and monounsaturated acyl-ACPs are direct substrates for the plastidial glycerol-3-phosphate acyltransferase and the lysophosphatidic acid acyltransferase, which catalyze the esterification of glycerol-3-phosphate at the sn-1 and sn-2 position. The resulting phosphatidic acid is the precursor for plastidial lipids, in which further desaturation of the acyl-residues can occur.

In the eukaryotic lipid biosynthesis pathway thioesterases cleave the fatty acids from the ACP cofactor and free fatty acids are exported to the cytoplasm where they participate as fatty acyl-CoA esters in the eukaryotic pathway. In this pathway the fatty acids are esterified by glycerol-3-phosphate acyltransferase and lysophosphatidic acid acyl-transferase to the sn-1 and sn-2 positions of glycerol-3-phosphate, respectively, to yield phosphatidic acid (PA). The PA is the precursor for other polar and neutral lipids, the latter being formed in the Kennedy ot other pathways (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).

The acyl-CoAs resulted from the export of plastidic fatty acids can also be elongated to yield very-long-chain fatty acids with more than 18 carbon atoms. Fatty acid elongases are multienzyme complexes consisting of at least four enzyme activities: beta-ketoacyl-CoA

synthases, beta-ketoacyl-CoA reductase, beta-hydroxyacyl-CoA dehydratase and enoyl-CoA reductase. It is well known that the beta-ketoacyl-CoA synthase determines the activity and the substrate selectivity of the fatty acid elongase complex (Millar & Kunst 1997, Plant J. 12:121-131). The very-long-chain fatty acids can be either used for wax and sphingolipid biosynthesis or enter the pathways for seed storage lipid biosynthesis.

Storage lipids in seeds are synthesized from carbohydrate-derived precursors. Plants have a complete glycolytic pathway in the cytosol (Plaxton 1996, Annu. Rev. Plant Physiol. Plant Mol. Biol. 47:185-214), and it has been shown that a complete pathway also exists in the plastids of rapeseeds (Kang & Rawsthorne 1994, Plant J. 6:795-805). Sucrose is the primary source of carbon and energy, transported from the leaves into the developing seeds. During the storage phase of seeds, sucrose is converted in the cytosol to provide the metabolic precursors glucose-6-phosphate and pyruvate. These are transported into the plastids and converted into acetyl-CoA that serves as the primary precursor for the synthesis of fatty acids. Acetyl-CoA in the plastids is the central precursor for lipid biosynthesis. Acetyl-CoA can be formed in the plastids by different reactions and the exact contribution of each reaction is still being debated (Ohlrogge & Browse 1995, Plant Cell 7:957-970). It is however accepted that a large part of the acetyl-CoA is derived from glucose-6-phospate and pyruvate that are imported from the cytoplasm into the plastids. Sucrose is produced in the source organs (leaves, or anywhere where 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.

Generally the breakdown of lipids is considered to be performed in plants in peroxisomes in the process know as beta-oxidation. This proecess involves the enzymatic reactions of acyl-CoA oxidase, hydroxyacyl-CoA-dehydrogenase (both found as a multifunctional complex) and ketoacyl-CoA-thiolase, with catalase in a supporting role (Graham and Eastmond 2002). In addition to the breakdown of common fatty acids beta-oxidation also plays a role in the removal of unusual fatty acids and fatty acid oxidation products, the glyoxylate cycle and the metabolism of branched chain amino acids (Graham and Eastmond 2002).

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

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 Δ12-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 Δ6-desaturase nucleic acid, Δ12-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.

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

Although several compounds are known that generally affect plant and seed development, there is a clear need to specifically identify factors, and particularly combinations thereof, that are more specific for the developmental regulation of storage compound accumulation and to identify combination of genes which have the capacity to confer altered or increased oil

production to its host plant and to other plant species. This invention discloses combinations of nucleic acid sequences from Phycomitrella patens and Arabidopsis thaliana. These combinations of 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, and peanut, which are oilseed plants containing high amounts of lipid compounds.

SUMMARY OF THE INVENTION

The present invention provides novel combinations isolated nucleic acid coding for LMPs and order thereof within the combinations, resulting in coordinated presence of proteins associated with the metabolism of seed storage compounds in plants

Also provided by the present invention are regulatory genetic elements such as promoters and terminators particularly suited for the exspression of combinations of more than one LMPs.

Also provided in the present invention is an arrangement of regulatory elements and genes encoding for LMPs to enhance their effect on seed storage compounds.

A further object of the present invention is an isolated nucleic acid comprising two or more LMP polynucleotide sequences selected from the group consisting of:

a. a polynucleotide sequence as described by SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13 or SEQ ID NO: 15;
b. a polynucleotide sequence encoding a polypeptide as described by SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 or SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14 or SEQ ID NO: 16;
c. a polynucleotide sequence having at least 70% sequence identity with the nucleic acid of a) or b) above;
d. a polynucleotide sequence that is complementary to the nucleic acid of a) or b) above; and
e. a polynucleotide sequence that hybridizes under stringent conditions to nucleic acid of a) or b) above.

Preferably, the isolated nucleic acid of the present invention encodes a polypeptide that functions as a modulator of a seed storage compound in microorganisms or plants. The nucleic acid of the present invention can comprise one, two, three, four, five, six, seven or eight of the nucleotide sequences of the present invention, preferably of the nucleotide sequences as described by SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13 or SEQ ID NO: 15. Especially preferred LMP nucleic acid sequences

are shown in the following table, wherein the sequence identifier are those used in WIPO Standard ST. 25 sequence listing.

Table 1: Combination of LMP nucleotide sequences

Combination of nucleotide sequences	Nucleotide sequence as described by SEQ ID NO:	Nucleotide sequence as described by SEQ ID NO:	Nucleotide sequence as described by SEQ ID NO:
21	1	11	
22	1	9	
23	7	11	
24	1	15	
25	1	3	
26	3	7	
27	3	9	
28	5	13	
29	1	9	13
30	1	7	3
31	9	15	13
32	5	3	
33	1	5	13
34	3	15	13
35	9	5	13
36	9	7	13
37	7	9	13
38	7	3	13
39	7	11	13
40	9	11	13
41	3	11	13
42	3	5	

Especially preferred are combinations number 21, 23, 26, 27, 32 & 33 of table 1. Further preferred nucleic acid sequences are the combinations of polynucleotide sequences shown in Figure 8, Table 9Table 9. Especially preferred are combinations number 21, 23, 26, 27, 32 & 33 of Figure 8, Table 9Table 9. The nucleic acids of the present invention, particularly those combinations of polynucleotide sequences shown in Figure 8, Table 9Table 9, further preferred

combinations number 21, 23, 26, 27, 32 & 33 of Figure 8, Table 9Table 9 can be used to increase the seed oil content in seeds , by e.g. 1, 2.5, 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5, 25, 27.5 or 30 % by weight or more, preferably by 5 % by weight or more, more preferably by 7.5 % by weight or more and even more preferably by 10 % by weight or more as compared to an empty vector control. Preferably the modification of the level or composition of a seed storage compound is measured as dry weight as measured by gas chromatography in the seed.

A further object is an isolated polypeptide encoded by a polynucleotide sequence above. Preferably the isolated polypeptide sequence of the present invention functions as a modulator of a seed storage compound in microorganisms or plants.
A further object of the present invention is an expression vector containing the nucleic acid of the present invention, wherein the nucleic acid is operatively linked to a promoter selected from the group consisting of a seed-specific promoter, a root-specific promoter, a leaf specific promoter and a non-tissue-specific promoter. Preferably the expression vector contains the combinations of polynucleotide sequences shown in Figure 8, Table 9. Especially preferred are combinations number 21, 23, 26, 27, 32 & 33 of Figure 8, Table 9.

By promoter is meant a polynucleotide sequence upstream from the transcriptional initiation site and which contains the regulatory regions required for transcription.
Preferably the promoter of the present invention is selected from the group consisting of:

a. a polynucleotide sequence as described by SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12 or SEQ ID NO: 13;
b. a polynucleotide sequence having at least 70% sequence identity with the nucleic acid of a) above;
c. a polynucleotide sequence that hybridizes under stringent conditions to nucleic acid of a) or b) above;
d. a polynucleotide sequence comprising at least 50 %, preferably 60%, 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% by number of polynucleotide sequences as described by the sequence identifiers (SEQ ID NO:) in the sequence listing of the present application of the polynucleotide sequences as described by the capital letters (e.g. ACAC for the polynucleotide sequence as described by SEQ ID NO: 26) of the polynucleotide sequence as described by SEQ ID NO: 26-156 related to a promoter as described by columns 1 and 2 of table 10 (e.g. for the promoter as described by SEQ ID NO: 9 at least 50 % of the polynucleotide sequences as described by SEQ ID NO: 44, 45, 46, 46, 48, 54, 59, 59, 59, 62, 63, 68, 70, 80, 80, 80, 81, 84, 84, 85, 87, 96, 97, 100, 105, 106, 108, 108, 108, 109,

114, 115, 115, 124, 124, 125, 135, 135, 136, 136, 141, 141, 142, 142, 142, 142, 144, 146, 146, 146, 148, 149, 152, 154, 154, 154. That means for the promoter as described by SEQ ID NO: 9 at least 50 % of the polynucleotide sequences as described by column 2, lines 2 to 57, preferably for the the promoter as described by SEQ ID NO: 10 at least 50 % of the polynucleotide sequences as described by column 2, lines 58 to 134. A polynucleotide sequence as described by SEQ ID NO: 26-156 can occur one or more times, e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more times in the promoter of the present invention, preferably as shown by the repetitions of polynucleotide sequences as described by the sequence identifiers in column 2 of table 10 of the promoter sequences as described by SEQ ID NO: 9-13); and

e. a polynucleotide sequence comprising a polynucleotide sequence having at least 70%, preferably 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% sequence identity with the full length polynucleotide sequence as described by the capital letters of the polynucleotide sequence as described by SEQ ID NO: 26-156, wherein the polynucleotide sequence comprises 50%, preferably 60%, 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% by number of polynucleotide sequences of the nucleotide sequences having at least 70% preferably 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% sequence identity with the polynucleotide sequence as described by the capital letters of the polynucleotide sequence as described by SEQ ID NO: 26-156 related to a promoter as described by columns 1 and 2 of table 10. The percent sequence identity between two polynucleotide sequences that are comprised in a promoter of the present invention is determined as the so called Core Similarity using the function MatInspector with default parameters of GEMS Launcher 4.2.2 software package (Genomatix Software GmbH). The algorithm underlying the Core Similarity is disclosed on pages 4879 - 4880 of Quandt K, Frech K, Karas H, Wingender E, Werner T (1995), MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Res. 23, 4878-4884, [PUBMED: 96128303])

In a preferred embodiment the promoter of the present invention comprises at least 50 %, preferably 60%, 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% by number of polynucleotide sequences of the full-length polynucleotide sequences as described by SEQ ID NO: 26-156 related to a promoter as described by columns 1 and 2 of table 10.

In a further preferred embodiment the promoter of the present invention comprises a polynucleotide sequence having at least 70%, preferably 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity with the full length polynucleotide sequence as described by SEQ ID NO: 26-156, wherein the polynucleotide sequence comprises 50%, preferably 60%, 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% by number of polynucleotide sequences as described by the sequence identifiers (SEQ ID NO:) related to a promoter as described by columns 1 and 2 of table 10 in the sequence listing of the present application of the nucleotide sequences having at least 70% sequence identity with the polynucleotide sequence as described by SEQ ID NO: 26-156. The percent sequence identity between two polynucleotide sequences that are comprised in a promoter of the present invention is determined as the so called Matrix Similarity using the function MatInspector with default parameters of GEMS Launcher 4.2.2 software package (Genomatix Software GmbH). The algorithm underlying the Matrix Similarity is disclosed on pages 4879 - 4880 of Quandt K, Frech K, Karas H, Wingender E, Werner T (1995), MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Res. 23, 4878-4884, [PUBMED: 96128303].

In a further preferred embodiment the promoter of the present invention comprises at least 50 %, preferably 60%, 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% by number of polynucleotide sequences of the polynucleotide sequences as described by the capital letters of the polynucleotide sequence as described by SEQ ID NO: 157-631 related to a promoter as described by columns 1 and 3 of table 10. For example for the promoter as described by SEQ ID NO: 9 at least 50 %, preferably 60%, 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% by number of polynucleotide sequences as described by the sequence identifiers (SEQ ID NO:) in the sequence listing of the present application of the polynucleotide sequences as described by SEQ ID NO: 263, 285, 303, 304, 313, 301, 260, 268, 271, 265, 264, 279, 277, 282, 294, 305, 290, 308, 310, 270, 278, 281, 262, 289, 300, 292, 275, 283, 287, 296, 293, 280, 286, 261, 314, 298, 272, 291, 307, 312, 297, 311, 276, 295, 302, 306, 267, 269, 274, 309 are comprised.

In a further preferred embodiment the promoter of the present invention comprises a polynucleotide sequence having at least 70%, preferably 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity with the polynucleotide sequences as described by the capital letters of the

polynucleotide sequence as described by SEQ ID NO: 157-631, wherein the polynucleotide sequence comprises 50%, preferably 60%, 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% by number of the polynucleotide sequences related to a promoter as described by columns 1 and 3 of table 10 having at least 70% sequence identity with the polynucleotide sequence as described by the capital letters of the polynucleotide sequence as described by SEQ ID NO: 157-631.

In a further preferred embodiment the promoter comprises at least 50 %, preferably 60%, 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% by number of polynucleotide sequences of the full-length polynucleotide sequences as described by SEQ ID NO: 157-631 related to a promoter as described by columns 1 and 3 of table 10.

In a further preferred embodiment the promoter comprises a polynucleotide sequence having at least 70%, preferably 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity with the polynucleotide sequences as described by SEQ ID NO: 157-631, wherein the polynucleotide sequence comprises 50%, preferably 60%, 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% by number of the polynucleotide sequences related to a promoter as described by columns 1 and 3 of table 10 having at least 70% preferably 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity with the polynucleotide sequence as described by SEQ ID NO: 157-631,

Preferably the expression vector of the present invention contains a terminator selected from the group consisting of:

a. a polynucleotide sequence as described by SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24 or SEQ ID NO: 25;
b. a polynucleotide sequence having at least 70% sequence identity with the nucleic acid of a) above;
c. a polynucleotide sequence that is complementary to the nucleic acid of a) or b) above; and
d. a polynucleotide sequence that hybridizes under stringent conditions to nucleic acid of a) or b) above.

A further object of the present invention is a method of producing a transgenic plant having a modified level of a seed storage compound weight percentage compared to an empty vector control comprising

a. a first step of introduction into a plant cell of an expression vector containing a nucleic acid, and
b. a further step of generating from the plant cell the transgenic plant,

wherein the nucleic acid encodes a polypeptide that functions as a modulator of a seed storage compound in the plant, and wherein the nucleic acid comprises a polynucleotide sequence of the present invention. In a preferred embodiment of the method of the present invention the nucleic acid comprises a polynucleotide sequence having at least 90% sequence identity with the polynucleotide sequence of the present invention, preferably the combinations of polynucleotide sequences shown in Figure 8, Table 9. Especially preferred are combinations number 21, 23, 26, 27, 32 & 33 of Figure 8, Table 9.

Preferably the total seed oil content weight percentage is increased in the transgenic plant as compared to an empty vector control, by e.g. 1, 2.5, 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5, 25, 27.5 or 30 % by weight or more, preferably by 5 % by weight or more, more preferably by 7.5 % by weight or more and even more preferably by 10 % by weight or more as compared to an empty vector control. Preferably for the purposes of this invention the modification of the level or composition of a seed storage compound is measured as dry weight as measured by gas chromatography in the seed.

The percent increases of a seed storage compound are generally 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 the present invention except for such a vector or construct lacking the nucleic acid sequences of the present inventions, preferably the nucleic acid sequences as disclosed in Appendix A. An empty vector control is shown for example in example 9.

A further object of the present invention is a method of modulating the level of a seed storage compound weight percentage in a plant comprising, modifying the expression of a nucleic acid in the plant, comprising

a. a first step of introduction into a plant cell of an expression vector comprising a nucleic acid, and
b. a further step of generating from the plant cell the transgenic plant,

wherein the nucleic acid encodes a polypeptide that functions as a modulator of a seed storage compound in the plant wherein the nucleic acid comprises the polynucleotide sequence of the present invention, preferably the combinations of polynucleotide sequences shown in Figure 8, Table 9. Especially preferred are combinations number 21, 23, 26, 27, 32 & 33 of Figure 8, Table 9.

The method of Claim 11, wherein the total seed oil content weight percentage is increased in the transgenic plant as compared to an empty vector control, , by e.g. 1, 2.5, 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5, 25, 27.5 or 30 % by weight or more, preferably by 5 % by weight or more, more preferably by 7.5 % by weight or more and even more preferably by 10 % by weight or more as compared to an empty vector control.

A further object of the present invention is a transgenic plant made by the method of the present invention. The transgenic plant of Claim 13, wherein the total seed oil content weight percentage is increased in the transgenic plant as compared to an empty vector control, by e.g. 1, 2.5, 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5, 25, 27.5 or 30 % by weight or more, preferably by 5 % by weight or more, more preferably by 7.5 % by weight or more and even more preferably by 10 % by weight or more as compared to an empty vector control. Preferably the transgenic plant of the present invention is selected from the group consisting of rapeseed, canola, linseed, soybean, sunflower, maize, oat, rye, barley, wheat, rice, pepper, tagetes, cotton, oil palm, coconut palm, flax, castor, sugarbeet, rice and peanut.
A further object of the present invention is a seed produced by the transgenic plant of the present invention, wherein the plant expresses the polypeptide that functions as a modulator of a seed storage compound and wherein the plant is true breeding for a modified level of seed storage compound weight percentage as compared to an empty vector control. The modification can be an increase or a decrease, preferably an increase of the seed storage compound, preferably of the seed oil content by e.g. 1, 2.5, 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5, 25, 27.5 or 30 % by weight or more, preferably by 5 % by weight or more, more preferably by 7.5 % by weight or more and even more preferably by 10 % by weight or more as compared to an empty vector control. Preferably the modification of the level or composition of a seed storage compound is measured as dry weight as measured by gas chromatography in the seed.

Additionally, the present invention relates to and provides the use of combinations LMP nucleic acids in the production of transgenic plants having a modified level or composition of a seed storage compound, preferably of seed oil, by e.g. 1, 2.5, 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5, 25, 27.5 or 30 % by weight or more, preferably by 5 % by weight or more, more preferably

by 7.5 % by weight or more and even more preferably by 10 % by weight or more as compared to an empty vector control.

For the purposes of the present invention, preferably the modification of the level or composition of a seed storage compound is measured as dry weight by gas chromatography in the seed.

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, by e.g. 1, 2.5, 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5, 25, 27.5 or 30 % by weight or more, preferably by 5 % by weight or more, more preferably by 7.5 % by weight or more and even more preferably by 10 % by weight or more as compared to an empty vector control. 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 an LMP nucleic acid, and generating a plant with a modified level or composition of the seed storage compound from the plant cell. In a preferred embodiment, the plant is an oil producing species selected from the group consisting of canola, linseed, soybean, sunflower, maize, oat, rye, barley, wheat, rice, pepper, tagetes, cotton, oil palm, coconut palm, flax, castor, and peanut, for example.

According to the present invention, the compositions and methods described herein can be used to alter the composition of more than one LMP in a transgenic plant and to increase or decrease the level of more than one LMP in a transgenic plant comprising increasing or decreasing the expression of more than one LMP nucleic acid in the plant, by e.g. 1, 2.5, 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5, 25, 27.5 or 30 % by weight or more, preferably by 5 % by weight or more, more preferably by 7.5 % by weight or more and even more preferably by 10 % by weight or more as compared to an empty vector control. Increased or decreased expression of the LMP nucleic acid can be achieved through transgenic overexpression, cosuppression approaches, antisense approaches, and in vivo mutagenesis of the LMP nucleic acid. The present invention can also be used to increase or decrease the level of a lipid in a seed oil, to increase or decrease the level of a fatty acid in a seed oil, or to increase or decrease the level of a starch in a seed or plant, by e.g. 1, 2.5, 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5, 25, 27.5 or 30 % by weight or more, preferably by 5 % by weight or more, more preferably by 7.5 % by weight or more and even more preferably by 10 % by weight or more as compared to an empty vector control.

More specifically, the present invention includes and provides a method for increasing total oil content in a seeds, by e.g. 1, 2.5, 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5, 25, 27.5 or 30 % by weight or more, preferably by 5 % by weight or more, more preferably by 7.5 % by weight or more and even more preferably by 10 % by weight or more as compared to an empty vector control comprising: transforming a plant with a nucleic acid construct that comprises as

operably linked components, combinations of two or more promoters and nucleic acid sequences encoding for LMPs, and growing the plant. Furthermore, the present invention includes and provides a method for increasing the level of oleic acid in a seed, by e.g. 1, 2.5, 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5, 25, 27.5 or 30 % by weight or more, preferably by 5 % by weight or more, more preferably by 7.5 % by weight or more and even more preferably by 10 % by weight or more as compared to an empty vector controlcomprising: transforming a plant with a nucleic acid construct that comprises as operably linked components, combinations of two or more promoters and nucleic acid sequences capable of increasing the level of oleic acid, and growing the plant.

Also included herein is a seed produced by a transgenic plant transformed by a combination of LMP DNA sequences, wherein the seed contains the LMP DNA sequences in a combination as described within this invention and wherein the plant is true breeding for a modified level of a seed storage compound. The present invention additionally includes a seed oil produced by the aforementioned seed.

Further provided by the present invention are vectors comprising the nucleic acids and combinations of the later, 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, by e.g. 1, 2.5, 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5, 25, 27.5 or 30 % by weight or more, preferably by 5 % by weight or more, more preferably by 7.5 % by weight or more and even more preferably by 10 % by weight or more as compared to an empty vector control.

It is further provided a method of producing a higher or lower, by e.g. 1, 2.5, 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5, 25, 27.5 or 30 % by weight or more, preferably by 5 % by weight or more, more preferably by 7.5 % by weight or more and even more preferably by 10 % by weight or more as compared to an empty vector control, than normal or typical level of storage compound in a transgenic plant expressing a combination of LMP nucleic acids in the transgenic plant, wherein the transgenic plant is Arabidopsis thaliana, Brassica napus, Glycine max, Oryza sativa, Zea mays, Triticum aestivum, Helianthus anuus or Beta vulgaris or a species different from Arabidopsis thaliana, Brassica napus, Glycine max, Oryza sativa or Triticum aestivum. 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 anuus 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 anuus and/or Beta vulgaris.

Accordingly, it is an object of the present invention to provide novel combinations of LMP nucleic acids and resulting in coordinate prduction of LMP amino acid sequences, 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.

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, by e.g. 1, 2.5, 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5, 25, 27.5 or 30 % by weight or more, preferably by 5 % by weight or more, more preferably by 7.5 % by weight or more and even more preferably by 10 % by weight or more as compared to an empty vector control.

The polynucleotides and combinations of the later of the present invention, including agonists and/or fragments thereof and of their encoded amino acid sequences, 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.

The combination of nucleic acid molecules of the present invention may further comprise combinations of operably linked promoter or partial promoter region. The promoters 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 (PF 55368-2 US, Song H. et al., 2004, see Example 11). 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 seed-specific promoter can be, for example, the USP promoter (Bäumlein et al. 1991, Mol. Gen. Genetics 225:459-67) or the leguminB4 promoter (Bäumlein et al. 1992, Plant Journal 2(2): 233-238). The tissue-specific promoter active in vegetative tissue can be a root-specific, shoot-

specific, meristem-specific or leaf-specific promoter. The isolated nucleic acid molecule of the present invention can still further comprise a 5' non-translated sequence, 3' non-translated sequence, introns, or the combination thereof.

The present invention also provides a method for increasing the number and/or size of one or more plant organs of a plant expressing a combination of nucleic acids encoding Lipid Metabolism Proteins (LMP), or a portion thereof, by e.g. 1, 2.5, 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5, 25, 27.5 or 30 % by weight or more, preferably by 5 % by weight or more, more preferably by 7.5 % by weight or more and even more preferably by 10 % by weight or more as compared to an empty vector control. More specifically, seed size and/or seed number and/or weight might be manipulated. Moreover, root length can be increased, by e.g. 1, 2.5, 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5, 25, 27.5 or 30 % by weight or more, preferably by 5 % by weight or more, more preferably by 7.5 % by weight or more and even more preferably by 10 % by weight or more as compared to an empty vector control. Longer roots can alleviate not only the effects of water depletion from soil but also improve plant anchorage/standability thus 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 lightcapture capacity and photosynthetic efficiency, by e.g. 1, 2.5, 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5, 25, 27.5 or 30 % by weight or more, preferably by 5 % by weight or more, more preferably by 7.5 % by weight or more and even more preferably by 10 % by weight or more as compared to an empty vector control.

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

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

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

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-H. SEQ ID NO:1-8 - open reading frame of the nucleic acid sequence coding for LMP useful in novel combinations to increase seed storage compounds.

Figures 2A-E. SEQ ID NO:9-13 Nucleic acid sequences of exemplary promoters useful in novel combinations to increase seed storage compounds.

Figures 3A-3B. SEQ ID NO:5-614-17 - Nucleic acid sequences of exemplary terminators useful in novel combinations to increase seed storage compounds.

Figure 4: Schematic representation of the binary vector pSUN indicating relevant features and restriction sites. b-RB = right border of T-DNA; c-aadA = aminoglycoside 3'-adenylyl-transferase codons; o-ColE1 replication origin of the plasmid pBR322, consisiting of the two components o-REP-ColE1 and o-BOM-ColE1; VS1-rep = replication origin and repA of plasmid pVS1 VS1-sta = sta gene from plasmid pVS1; b-LB = left border of T-DNA; T-DNA cassette marks the region where the different T-DNA cassette for the different constructs are located. These T-DNA cassettes are described in Figure 5.

Figure 5: Schematic representation of the T-DNA cassette containing the arrangement of the novel combination of genes coding for LMPs. Positions within the combination are given by the letters A-C, SM denotes the selection marker cassette elements (promoter, selection marker gene & terminator); b-LB = left border of T-DNA, b-RB = right border of T-DNA

Figure 6: Graphical representation of the seed oil content of Arabidopsis T2 seeds carrying combinations of LMPs number 21, 23, 26, 27, 32 & 33 (see table 9 of Figure 8). Graphs represent the g fatty acids in the seed per g dry weight as measured by gas chromatography. Black bars represent lines carrying the combinations, empty bars represent the values from 3 empty vector controls. Each value is the mean of at least duplicate extractions and measurements, error bars represent the standard deviation. The control value given is the mean of 3 to 8 empty vector controls extracted and measured in at least duplicate.
Table 7 of figure 6 provides the peak relative increase in seed oil content of T2 Arabidopsis seed harboring the combinations of LMPS as measured by gas chromatography as described above.

As a further example the data shown in table 8 in figure 7 demonstrates that seed oil content of canola seed can significantly be increased by introduction of the combinations of LMPs as listed in table 9 of figure 8. T2 seeds of plants harbouring the combination of LMPs listed in table 8 were analysed for seed oil content by NIRS. Control plants were non-transgenic segregants grown together with the transgenic plants carrying the combination of LMPs.Only lines with an increase of more than 5 % are shown. The p-values shown were calculated using simple t-test.

DETAILED DESCRIPTION OF THE INVENTION

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

Before the present compounds, compositions, and methods are disclosed and described, it is to be understood that this invention is not limited to specific nucleic acids, specific

polypeptides, specific cell types, specific host cells, specific conditions, or specific methods, etc., as such may, of course, vary, and the numerous modifications and variations therein will be apparent to those skilled in the art. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting. As used in the specification and in the claims, “a” or “an” can mean one or more, depending upon the context in which it is used. Thus, for example, reference to “a cell” can mean that at least one cell can be utilized.

The term "transgenic" or "recombinant" when used in reference to a cell or an organism (e.g., with regard to a barley plant or plant cell) refers to a cell or organism which contains a transgene, or whose genome has been altered by the introduction of a transgene. 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 by man. Preferably, said sequence is resulting in a genome which is different from a naturally occurring 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 be an "endogenous DNA sequence", “an “exogenous DNA sequence” (e.g., a foreign gene), or a "heterologous DNA sequence". The term "endogenous DNA sequence" refers to a nucleotide sequence, which is naturally found in the cell into which it is introduced so long as it does not contain some modification (e.g., a point mutation, the presence of a selectable marker gene, etc.) relative to the naturally-occurring sequence.

The term "wild-type", "natural" or of "natural origin" means with respect to an organism, polypeptide, or nucleic acid sequence, that said organism is naturally occurring or available in at least one naturally occurring organism which is not changed, mutated, or otherwise manipulated by man.

The terms "heterologous nucleic acid sequence" or "heterologous DNA" are used interchangeably to refer to a nucleotide sequence, which is ligated to, or is manipulated to become ligated to, a nucleic acid sequence to which it is not ligated in nature, or to which it is ligated at a different location in nature. Heterologous DNA is not endogenous to the cell into which it is introduced, but has been obtained from another cell. Generally, although not necessarily, such heterologous DNA encodes RNA and proteins that are not normally produced by the cell into which it is expressed. A promoter, transcription regulating sequence or other genetic element is considered to be “heterologous” in relation to another sequence (e.g., encoding a marker sequence or am agronomically relevant trait) if said two sequences are not

combined or differently operably linked their natural environment. Preferably, said sequences are not operably linked in their natural environment (i.e. come from different genes). Most preferably, said regulatory sequence is covalently joined and adjacent to a nucleic acid to which it is not adjacent in its natural environment.

One aspect of the invention pertains to combinations of 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. 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 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. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule consisting of a combination of isolated nucleotide sequences of Appendix A, or a portion thereof, can be constructed using standard molecular biology techniques and the sequence information provided herein. For example, an Arabidopsis thaliana or Physcomitrella patens, Brassica napus, Glycine max or Linum usitatissimum LMP cDNA can be isolated from an Arabidopsis thaliana or Physcomitrella patens, , Brassica napus, Glycine max or Linum usitatissimum library using all or portion of one of the sequences of Appendix A as a hybridization probe and standard hybridization techniques (e.g., as described in Sambrook et al. 1989, Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). Moreover, a nucleic acid molecule encompassing

all or a portion of one of the sequences of Appendix A can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this sequence (e.g., a nucleic acid molecule encompassing all or a portion of one of the sequences of Appendix A can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this same sequence of Appendix A). For example, mRNA can be isolated from plant cells (e.g., by the guanidinium-thiocyanate extraction procedure of Chirgwin et al. 1979, Biochemistry 18:5294-5299) and cDNA can be prepared using reverse transcriptase (e.g., Moloney MLV reverse transcriptase, available from Gibco/BRL, Bethesda, MD; or AMV reverse transcriptase, available from Seikagaku America, Inc., St. Petersburg, FL). Synthetic oligonucleotide primers for polymerase chain reaction amplification can be designed based upon one of the nucleotide sequences shown in Appendix A and may contain restriction enzyme sites or sites for ligase independent cloning to construct the combinations described by this invention. 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 acids so amplified can be cloned into an appropriate vector in the combinations described by the present invention or variations thereof and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to an LMP nucleotide sequence can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

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

In still another preferred embodiment, an isolated nucleic acid molecule in the combinations 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 one or more nucleotide sequence shown in Appendix A, or a portion thereof. In an additional preferred embodiment, an isolated nucleic acid molecule in the combinations of the invention comprises a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to one or more of the nucleotide sequences shown in Appendix A, or a portion thereof.

For the purposes of the invention hybridzation means preferably hybridization under conditions equivalent to hybridization in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO4, 1 mM EDTA at 50°C with washing in 2 X SSC, 0. 1% SDS at 50°C, more desirably in 7% sodium

dodecyl sulfate (SDS), 0.5 M NaPO4, 1 mM EDTA at 50°C with washing in 1 X SSC, 0.1% SDS at 50°C, more desirably still in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO4, 1 mM EDTA at 50°C with washing in 0.5 X SSC, 0. 1% SDS at 50°C, preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO4, 1 mM EDTA at 50°C with washing in 0.1 X SSC, 0.1% SDS at 50°C, more preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO4, 1 mM EDTA at 50°C with washing in 0.1 X SSC, 0.1% SDS at 65°C to a nucleic acid comprising 50 to 200 or more consecutive nucleotides.

A further preferred, non-limiting example of stringent hybridization conditions includes washing with a solution having a salt concentration of about 0.02 molar at pH 7 at about 60°C.

Moreover, the nucleic acid molecule in the combinations of the invention can comprise only a portion of the coding region of one of the sequences in Appendix A, for example a fragment, which can be used as a probe or primer or a fragment encoding a biologically active portion of an LMP. The nucleotide sequences determined from the cloning of the LMP Arabidopsis thaliana or Physcomitrella patens, 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 combinations of nucleic acids disclosed herein, or fragments thereof. These compounds include the nucleic acid combinations attached to a moiety. These moieties include, but are not limited to, detection moieties, hybridization moieties, purification moieties, delivery moieties, reaction moieties, binding moieties, and the like. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 40, 50, or 75 consecutive nucleotides of a sense strand of one of the sequences set forth in Appendix A, an anti-sense sequence of one of the sequences set forth in Appendix A, or naturally occurring mutants thereof. Primers based on a nucleotide sequence of Appendix A can be used in PCR reactions to clone LMP homologues for the combinations described by this inventions or variations thereof. 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 an LMP, such as by measuring a level of an 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 combination of proteins or portions thereof, which include amino acid sequences, which are sufficiently

homologous to an amino acid encoded by a sequence of Appendix A, such that the protein or portion thereof maintains the same or a similar function as the wild-type protein. As used herein, the language "sufficiently homologous" refers to proteins or portions thereof, which have amino acid sequences, which include a minimum number of identical or equivalent (e.g., an amino acid residue, which has a similar side chain as an amino acid residue in one of the ORFs of a sequence of Appendix A) amino acid residues to an amino acid sequence, such that the protein or portion thereof is able to participate in the metabolism of compounds necessary for the production of seed storage compounds in plants, construction of cellular membranes in microorganisms or plants, or in the transport of molecules across these membranes. Examples of LMP-encoding nucleic acid sequences are set forth in Appendix A.

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 an LMP" is intended to include a portion, e.g., a domain/motif, of an LMP that participates in the metabolism of compounds necessary for the biosynthesis of seed storage lipids, or the construction of cellular membranes in microorganisms or plants, or in the transport of molecules across these membranes, or has an activity as set forth in Table 4. To determine whether an 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 an LMP include peptides comprising amino acid sequences derived from the amino acid sequence of an LMP (e.g., an amino acid sequence encoded by a nucleic acid of Appendix A or the amino acid sequence of a protein homologous to an LMP, which include fewer amino acids than a full length LMP or the full length protein which is homologous to an LMP) and exhibit at least one activity of an 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 an LMP. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the

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

Additional nucleic acid fragments encoding biologically active portions of an 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 combinations of nucleic acid molecules that differ from one of the nucleotide sequences shown in Appendix A (and portions thereof) due to degeneracy of the genetic code and thus encode the same LMP as that encoded by the nucleotide sequences shown in Appendix A. In a further embodiment, the combinations of nucleic acid molecule of the invention encode one or more full-length proteins, which are substantially homologous to an amino acid sequence of a polypeptide encoded by an open reading frame shown in Appendix A. In one embodiment, the full-length nucleic acid or protein, or fragment of the nucleic acid or protein, is from Arabidopsis thaliana or Physcomitrella patens.

In addition to the Arabidopsis thaliana or Physcomitrella patens LMP nucleotide sequences shown in Appendix A, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of LMPs may exist within a population Arabidopsis thaliana or Physcomitrella patens 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 an LMP, preferably an Arabidopsis thaliana or Physcomitrella patens 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.

The invention further encompasses combinations of nucleic acid molecules corresponding to natural variants and non- Arabidopsis thaliana or Physcomitrella patens orthologs of the Arabidopsis thaliana or Physcomitrella patens LMP nucleic acid sequence shown in Appendix A. Nucleic acid molecules corresponding to natural variants and non- Arabidopsis thaliana or Physcomitrella patens orthologs of the Arabidopsis thaliana or Physcomitrella patens LMP cDNA described in Appendix A can be isolated based on their homology to Arabidopsis thaliana or Physcomitrella patens LMP nucleic acid shown in Appendix A using the Arabidopsis thaliana or Physcomitrella patens cDNA, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. As used herein, the term “orthologs” refers to two nucleic acids from different species, but that have evolved from a common ancestral gene by speciation. Normally, orthologs encode proteins having the same or

similar functions. Accordingly, in another embodiment, an isolated nucleic acid molecule is at least 15 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising a nucleotide sequence of Appendix A. In other embodiments, the nucleic acid is at least 30, 50, 100, 250, or more nucleotides in length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing, under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 65%, more preferably at least about 70%, and even more preferably at least about 75% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y., 1989: 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. Preferably, an isolated nucleic acid molecule that hybridizes under stringent conditions to a sequence of Appendix A corresponds to a naturally occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

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

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding LMPs that contain changes in amino acid residues that are not essential for LMP activity. Such LMPs differ in amino acid sequence from a sequence yet retain at least one of the LMP activities described herein. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 50% homologous to an amino acid sequence encoded by a nucleic acid of Appendix A and is capable of participation in the metabolism of compounds necessary for the production of seed storage compounds in Brassica napus, Glycine max or Linum

usitatissimum, or cellular membranes, or has one or more activities set forth in Table 4. Preferably, the protein encoded by the nucleic acid molecule is at least about 50-60% homologous to one of the sequences encoded by a nucleic acid of Appendix A, more preferably at least about 60-70% homologous to one of the sequences encoded by a nucleic acid of Appendix A, even more preferably at least about 70-80%, 80-90%, 90-95% homologous to one of the sequences encoded by a nucleic acid of Appendix A, and most preferably at least about 96%, 97%, 98%, or 99% homologous to one of the sequences encoded by a nucleic acid of Appendix A.

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

For the purposes of the invention, the percent sequence identity between two nucleic acid or polypeptide sequences is 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.

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

more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into one of the sequences of Appendix A by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in an 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 an LMP coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for an LMP activity described herein to identify mutants that retain LMP activity. Following mutagenesis of one of the sequences of Appendix A, the encoded protein can be expressed recombinantly, and the activity of the protein can be determined using, for example, assays described herein (see Examples 11-13 of the Exemplification).

Combinations of LMPs are preferably produced by recombinant DNA techniques. For example, one or more 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 LMPs are expressed in the host cell. The LMPs can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Alternative to recombinant expression, one or more LMP or peptide thereof can be synthesized chemically using standard peptide synthesis techniques. Moreover, native LMPs can be isolated from cells, for example using an anti-LMP antibody, which can be produced by standard techniques utilizing an LMP or fragment thereof of this invention.

The invention also provides combinations of LMP chimeric or fusion proteins. As used herein, an LMP "chimeric protein" or "fusion protein" comprises an LMP polypeptide operatively linked to a non-LMP polypeptide. An "LMP polypeptide" refers to a polypeptide having an amino acid sequence corresponding to an LMP, whereas a "non-LMP polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the LMP, e.g., a protein which is different from the LMP, and which is derived from the same or a different organism. Within the fusion protein, the term "operatively

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

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

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

In another variation of the antisense technology, a double-strand, interfering, RNA construct can be used to cause a down-regulation of the LMP mRNA level and LMP activity in transgenic plants. This requires transforming the plants with a chimeric construct containing a portion of the LMP sequence in the sense orientation fused to the antisense sequence of the same portion

of the LMP sequence. A DNA linker region of variable length can be used to separate the sense and antisense fragments of LMP sequences in the construct.

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

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

In still another embodiment, a combination containing an antisense nucleic acid of the invention contains a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity, which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff & Gerlach 1988, Nature 334:585-591)) can be used to catalytically cleave LMP mRNA transcripts to thereby inhibit translation of LMP mRNA. A ribozyme having specificity for an LMP-encoding nucleic acid can be designed based upon the nucleotide sequence of an LMP cDNA disclosed herein (i.e., Bn01 in Appendix A) or on the basis of a heterologous sequence to be isolated according to methods taught in this invention. For example, a derivative of a Tetrahymena L-19 IVS RNA can be constructed, in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an 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).

Alternatively, LMP gene expression of one or more genes of the combinations of this invention can be inhibited by targeting nucleotide sequences complementary to the regulatory region of an LMP nucleotide sequence (e.g., an LMP promoter and/or enhancers) to form triple helical structures that prevent transcription of an 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 combination of nucleic acids encoding LMPs (or a portion thereof). As used herein, the term "vector" 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 (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes, to which they are operatively linked. Such vectors are referred to herein as "expression vectors." In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid," and "vector" can be used inter-changeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a combination of nucleic acids 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 Biotechnolgy, 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 combinations 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: an 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 purification 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 1988, Gene 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ), which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. In one embodiment, the coding sequence of the LMP is cloned into a pGEX expression vector to create a vector encoding a fusion protein comprising, from the N-terminus to the C-terminus, GST-thrombin cleavage site-X protein. The fusion protein can be purified by affinity chromatography using glutathione-agarose resin. Recombinant LMP unfused to GST can be recovered by cleavage of the fusion protein with thrombin.

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

One strategy to maximize recombinant protein expression is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman S. 1990, Gene Expression Technology: Methods in Enzymology 185:119-128, Academic Press, San Diego, California). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are

those preferentially utilized in the bacterium chosen for expression (Wada et al. 1992, Nucleic Acids Res. 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

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

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

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

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) (SEQ ID No. 16) or functional equivalents thereof. but also all other terminators functionally active in plants are suitable (e.g. Seq ID No. 14, 15 and 17).

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).

Plant gene expression has to be operably linked to an appropriate promoter conferring gene expression in a timely, cell or tissue specific manner. Preferred are promoters driving constitutive expression (Benfey et al. 1989, EMBO J. 8:2195-2202) like those derived from plant viruses like the 35S CAMV (Franck et al. 1980, Cell 21:285-294), the 19S CaMV (see also US 5,352,605 and WO 84/02913) or the ptxA promoter SEQ ID No. 9 (Bown,D.P. PhD thesis (1992) Department of Biological Sciences, University of Durham, Durham, U.K) or plant promoters like those from Rubisco small subunit described in US 4,962,028. Even more preferred are seed-specific promoters driving expression of LMP proteins during all or selected stages of seed development. Seed-specific plant promoters are known to those of ordinary skill in the art and are identified and characterized using seed-specific mRNA libraries and expression profiling techniques. Seed-specific promoters include the napin-gene promoter from rapeseed (US 5,608,152), the USP-promoter from Vicia faba (Baeumlein et al. 1991, Mol. Gen. Genetics 225:459-67) SEQ ID No. 10, 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) SEQ ID No. 11 & 12), as well as promoters conferring seed specific expression in monocot plants like maize, barley, wheat, rye, rice etc. Suitable promoters to note are the lpt2 or lpt1-gene promoter from barley (WO 95/15389 and WO 95/23230) or those described in WO 99/16890 (promoters from the barley hordein-gene, the rice glutelin gene, the rice oryzin gene, the rice prolamin gene, the wheat gliadin gene, wheat glutelin gene, the maize zein gene, the oat glutelin gene, the Sorghum kasirin-gene, and the rye secalin gene).

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. Examples for such promoters are a salicylic acid inducible promoter (WO 95/19443), a tetracycline inducible promoter (Gatz et al. 1992, Plant J. 2:397-404) and an ethanol inducible promoter (WO 93/21334).

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 acombination of DNA molecules 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 combination of LMPs 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 expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those that confer resistance to drugs, such as G418, hygromycin, kanamycin, and methotrexate or in plants that confer resistance towards an herbicide, such as glyphosate or glufosinate. A nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a combination of LMPs 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 that contains a combination of at least a portion of an 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 Physcomitrella patens LMP gene, but it can be a homologue from a related plant or even from a mammalian, yeast, or insect source. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous LMP gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a knock-out vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous LMP gene is mutated or otherwise altered

but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous LMP). To create a point mutation via homologous recombination, DNA-RNA hybrids can be used in a technique known as chimeraplasty (Cole-Strauss et al. 1999, Nucleic Acids Res. 27:1323-1330 and Kmiec 1999, American Scientist 87:240-247). Homologous recombination procedures in Arabidopsis thaliana or other crops are also well known in the art and are contemplated for use herein.

In a homologous recombination vector, within the combination of genes coding for LMPs shown in Appendix A 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 combinations of genes. For example, inclusion of a combination of one two or moreLMP genes on a vector placing it under control of the lac operon permits expression of the LMP gene only in the presence of IPTG. Such regulatory systems are well known in the art.

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 combination of LMPs. 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 combination of LMPs has been introduced, or which contains a wild-type or altered LMP gene in it's genome) in a suitable medium until the combination of LMPs is produced.

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, Glycine max or Linum usitatissimum or of cellular membranes, or has one or more of the activities set forth in Table 4. In preferred embodiments, the protein or portion thereof comprises an amino acid sequence which is sufficiently homologous to an amino acid sequence encoded by a nucleic acid of Appendix A such that the protein or portion thereof maintains the ability to participate in the metabolism of compounds necessary for the construction of cellular

membranes in Brassica napus, Glycine max or Linum usitatissimum, or in the transport of molecules across these membranes. The portion of the protein is preferably a biologically active portion as described herein. In another preferred embodiment, an LMP of the invention has an amino acid sequence encoded by a nucleic acid of Appendix A. In yet another preferred embodiment, the LMP has an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to a nucleotide sequence of Appendix A. In still another preferred embodiment, the LMP has an amino acid sequence which is encoded by a nucleotide sequence that is at least about 50-60%, preferably at least about 60-70%, more preferably at least about 70-80%, 80-90%, 90-95%, and even more preferably at least about 96%, 97%, 98%, 99%, or more homologous to one of the amino acid sequences encoded by a nucleic acid of Appendix A. The preferred LMPs of the present invention also preferably possess at least one of the LMP activities described herein. For example, a preferred LMP of the present invention includes an amino acid sequence encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to a nucleotide sequence of Appendix A, and which can participate in the metabolism of compounds necessary for the construction of cellular membranes in Brassica napus, Glycine max or Linum usitatissimum, or in the transport of molecules across these membranes, or which has one or more of the activities set forth in Table 4.

In other embodiments, the combination of LMPs is substantially homologous to a combination of amino acid sequences encoded by nucleic acids of Appendix A and retain the functional activity of the protein of one of the sequences encoded by a nucleic acid of Appendix A yet differs in amino acid sequence due to natural variation or mutagenesis, as described in detail above. Accordingly 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 or Physcomitrella patens, Brassica napus, Glycine max or Linum usitatissimum protein which is substantially homologous to an entire amino acid sequence encoded by a nucleic acid of Appendix A.

Dominant negative mutations or trans-dominant suppression can be used to reduce the activity of an 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 as part of the combination of this invention 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 for combinations 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 an LMP, which mediates transport of compounds across such membranes, thereby preventing translocation from taking place.

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 an LMP to be included in combinations as described in table 3. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of an 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 and fusion proteins for the combinations described herein, and vectors, and host cells described herein can be used in one or more of the following methods: identification of Arabidopsis thaliana or Physcomitrella patens and related organisms; mapping of genomes of organisms related to Arabidopsis thaliana or Physcomitrella patens; identification and localization of Arabidopsis thaliana or Physcomitrella patens sequences of interest; evolutionary studies; determination of LMP regions required for function; modulation of an 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.

The plant Arabidopsis thaliana represents one member of higher (or seed) plants. It is related to other plants such as Brassica napus, Glycine max or Linum usitatissimum which require light to drive photosynthesis and growth. Plants like Arabidopsis thaliana, Brassica napus, Glycine max or Linum usitatissimum 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, isolation of the corresponding genes and use of the later in combinations described for the sequences listed in Appendix A.

There are a number of mechanisms by which the alteration of a combination of LMPs of the invention may directly affect the accumulation and/or composition of seed storage compounds. In the case of plants expressing a combination of LMPs, increased transport can lead to altered accumulation of compounds, which ultimately could be used to affect the accumulation of one or more seed storage compounds during seed development. Expression of single genes affecting seed storage compound accumulation and/or solute partitioning within the plant tissue and organs is well known. 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 signaling pathway, and thereby in early and late seed development (e.g. Merlot et al. 2001, Plant J. 25:295-303). For more examples see also the section "Background of the Invention."

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.

It will be apparent to those skilled in the art that various modifications and variations can be made in the present invention without departing from the scope or spirit of the invention. Other embodiments of the invention will be apparent to those skilled in the art from consideration of the specification and practice of the invention disclosed herein. It is intended that the specification and Examples be considered as exemplary only, with a true scope and spirit of the invention being indicated by the claims included herein.

EXAMPLES

Example 1: General Processes - a) General Cloning Processes. 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).

Example 1: General Processes - 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 H2O 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), Roche (Mannheim), Genomed (Bad Oeynnhausen), 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.

Example 1: General Processes - c) Plant Material and Growth: Arabidopsis plants. For this study, root material, leaves, siliques and seeds of wild-type and transgenic plants of Arabidopsis thaliana expressing combinations of LMPs as described within this invention were used. Wild type and transgenic 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 µmol m-2 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.

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, 100µl of N-laurylsarcosine buffer, 20µl of β-mercaptoethanol and 10µl of proteinase K solution, 10 mg/ml) and incubated at 60°C for 1 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 180µl of TE buffer (Sambrook et al. 1989, Cold Spring Harbor Laboratory Press: ISBN 0-87969-309-6). For further purification, the DNA was treated with NaCl (1.2 M final concentration) and precipitated again at -70°C for 30 min using twice the volume of absolute ethanol. After a washing step with 70% ethanol, the DNA was dried and subsequently taken up in 50 µl of H2O + 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. 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:

1. Buffers, enzymes and solution

- 2M KCl
- Proteinase K
- Phenol (for RNA)
- Chloroform:Isoamylalcohol
  (Phenol:choloroform 1:1; pH adjusted for RNA)
- 4 M LiCl, DEPC-treated
- DEPC-treated water
- 3M NaOAc, pH 5, DEPC-treated
- Isopropanol
- 70% ethanol (made up with DEPC-treated water)
- Resuspension buffer:0.5% SDS, 10 mM Tris pH 7.5, 1 mM EDTA made up with DEPC-treated water as this solution cannot be DEPC-treated
- Extraction Buffer:

0.2M Na Borate
30 mM EDTA
30 mM EGTA
1% SDS (250µl of 10% SDS-solution for 2.5ml buffer)
1% Deoxycholate (25mg for 2,5ml buffer)
2% PVPP (insoluble - 50mg for 2.5ml buffer)
2% PVP 40K (50mg for 2.5ml buffer)
10 mM DTT
100 mM β-Mercaptoethanol (fresh, handle under fume hood - use 35µl of 14.3M solution 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. 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.

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 10 min at

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

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 chloform:isoamylalcohol extraction until interface is clear. Transfer supernatant to fresh tube and add 1/10 vol 3M NaOAc, pH 5 and 600µl isopropanol. Keep at -20 for 20 min or longer. Pellet RNA by 10 min centrifugation. Wash pellet once with 70% ethanol. Remove all remaining alcohol before resolving pellet with 15 to 20µl DEPC-water. Determine quantity and quality by measuring the absorbance of a 1:200 dilution at 260 and 280nm. 40µg RNA/ml = 1OD260

RNA from wild-type and the transgenic Arabidopsis-plants 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.

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

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.

Example 5: Northern-Hybridization. For RNA hybridization, 20µg of total RNA or 1µg 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.

Example 6: 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. 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´ in frame to the cDNA to achieve subcellular localization of the fusion protein.

Further examples for plant binary vectors are the pSUN300 or pSUN2-GW vectors, into which the combination of LMP genes are cloned. These binary vectors contain an antibiotic resistance gene driven under the control of the NOS promoter and combinations (see Table 9 of Figure 8) containing promoters as listed in Figure 2, LMP genes as shown in Figure 1 and terminators in Figure 3 Partial or full-length LMP cDNA are cloned into the multiple cloning site of the pEntry vector in sense or antisense orientation behind a seed-specific promoters or constitutive promoter (see Figure 2) in the combinations shown in Table 9 of Figure 8 using standard cloning procedures using restriction enzymes such as ASCI, PACI, NotP and StuI. Two or more pEntry vectors containing different LMPs are then combined with a pSUN destination vector to form a binary vector containing the combinations as listed in Table 9 of Figure 8 by the use of the GATEWAY technology (Invitrogen, http://www.invitrogen.com) following the manufacturer's instructions. The recombinant vector containing the combination of interest is transformed into Top10 cells (Invitrogen) using standard conditions. Transformed cells are selected for on LB agar containing 50µg/ml kanamycin grown overnight at 37°C.

Plasmid DNA is extracted using the QIAprep Spin Miniprep Kit (Qiagen) following manufacturer's instructions. Analysis of subsequent clones and restriction mapping is performed according to standard molecular biology techniques (Sambrook et al. 1989, Molecular Cloning, A Laboratory Manual. 2nd Edition. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY).

Example 7: Agrobacterium Mediated Plant Transformation. Agrobacterium mediated plant transformation with the combination of LMP nucleic acids described herein can be performed using standard transformation and regeneration techniques (Gelvin, Stanton B. & Schilperoort R.A, Plant Molecular Biology Manual, 2nd ed. Kluwer Academic Publ., Dordrecht 1995 in Sect., Ringbuc Zentrale Signatur:BT11-P; Glick, Bernard R. and Thompson, John E. Methods in Plant Molecular Biology and Biotechnology, S. 360, CRC Press, Boca Raton 1993). For example, Agrobacterium mediated transformation can be performed using the GV3 (pMP90) (Koncz & Schell, 1986, Mol. Gen. Genet. 204:383-396) or LBA4404 (Clontech) Agrobacterium tumefaciens strain.

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 combination of LMP nucleic acids of the present invention via cotyledon or hypocotyl transformation (Moloney et al. 1989, Plant Cell Report 8:238-242; De Block et al. 1989, Plant Physiol. 91:694-701). Use of antibiotic for Agrobacterium and plant selection depends on the binary vector and the Agrobacterium strain used for transformation. Rapeseed selection is normally performed using a selectable plant marker. Additionally, Agrobacterium mediated gene transfer to flax can be performed using, for example, a technique described by Mlynarova et al. (1994, Plant Cell Report 13:282-285).

The LMPs in the combinations described in this invention can be expressed either under the seed specific USP (unknown seed protein) promoter (Baeumlein et al. 1991, Mol. Gen. Genetics 225:459-67), the PtxA promoter (the promoter of the Pisum sativum PtxA gene), which is a promoter active in virtually all plant tissues or the superpromoter, which is a constitutive promoter (Stanton B. Gelvin, USP# 5,428,147 and USP#5,217,903) or other seed-specific promoters like 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.

The nptII gene was used as a selectable marker in these constructs. Figures 4 and 5 show the setup of the binary vectors containing the combinations of LMPs.

Transformation of soybean can be performed using, for example, a technique described in EP 0424 047, U.S. Patent No. 5,322,783 (Pioneer Hi-Bred International) or in EP 0397 687, U.S. Patent No. 5,376,543 or U.S. Patent No. 5,169,770 (University Toledo), or by any of a

number of other transformation procedures known in the art. Soybean seeds are surface sterilized with 70% ethanol for 4 minutes at room temperature with continuous shaking, followed by 20% (v/v) CLOROX supplemented with 0.05% (v/v) TWEEN for 20 minutes with continuous shaking. Then the seeds are rinsed 4 times with distilled water and placed on moistened sterile filter paper in a Petri dish at room temperature for 6 to 39 hours. The seed coats are peeled off, and cotyledons are detached from the embryo axis. The embryo axis is examined to make sure that the meristematic region is not damaged. The excised embryo axes are collected in a half-open sterile Petri dish and air-dried to a moisture content less than 20% (fresh weight) in a sealed Petri dish until further use.

The method of plant transformation is also applicable to Brassica napus and other crops. In particular, seeds of canola are surface sterilized with 70% ethanol for 4 minutes at room temperature with continuous shaking, followed by 20% (v/v) CLOROX supplemented with 0.05 % (v/v) TWEEN for 20 minutes, at room temperature with continuous shaking. Then, the seeds are rinsed four 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 resuspended 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 hours 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 500mg/l carbenicillin or 300mg/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 µmol m-2s-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 µmol m-2s-1 light intensity and 12-hour photoperiod for about 80 days.

Samples of the primary transgenic plants (T0) are analyzed by PCR to confirm the presence of T-DNA. These results are confirmed by Southern hybridization wherein DNA is electrophoresed on a 1% agarose gel and transferred to a positively charged nylon membrane (Roche Diagnostics). The PCR DIG Probe Synthesis Kit (Roche Diagnostics) is used to prepare a digoxigenin-labeled probe by PCR as recommended by the manufacturer.

Example 7: In vivo Mutagenesis. In vivo mutagenesis of microorganisms can be performed by incorporation and passage of the plasmid (or other vector) DNA through E. coli or other microorganisms (e.g. Bacillus spp. or yeasts such as Sacchromyces) 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. 2277-2294, ASM: Washington). Such strains are well known to those skilled in the art. The use of such strains is illustrated, for example, in Greener and Callahan 1994, Strategies 7:32-34. Transfer of mutated DNA molecules into plants is preferably done after selection and testing in microorganisms. Transgenic plants are generated according to various examples within the exemplification of this document.

Example 8: 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 measured on the transcriptional or/and on the translational level. A useful method to ascertain the level of transcription of the gene (an indicator of the amount of mRNA available for translation to the gene product) is to perform a Northern blot (for reference see, for example, Ausubel et al. 1988, Current Protocols in Molecular Biology, Wiley: New York), in which a primer designed to bind to the gene of interest is labeled with a detectable tag (usually radioactive or chemiluminescent), such that when the total RNA of a culture of the organism is extracted, run on gel, transferred to a stable matrix and incubated with this probe, the binding and quantity of binding of the probe indicates the presence and also the quantity of mRNA for this gene. This information at least partially demonstrates the degree of transcription of the transformed gene. Total cellular RNA can be prepared from plant cells, tissues or organs by several methods, all well-known in the art, such as that described in Bormann et al. (1992, Mol. Microbiol. 6:317-326).

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 8: In vitro Analysis of the activity of LMPS expressed in combinations 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 structure, kinetics, principles, methods, applications, and examples for the determination of many enzyme activities may be found, for example, in the following references: Dixon, M. & Webb, E.C. 1979, Enzymes. Longmans: London; Fersht, (1985) Enzyme Structure and Mechanism. Freeman: New York; Walsh (1979) Enzymatic Reaction Mechanisms. Freeman: San Francisco; Price, N.C., Stevens, L. (1982) Fundamentals of Enzymology. Oxford Univ. Press: Oxford; Boyer, P.D., ed. (1983) The Enzymes, 3rd ed. Academic Press: New York; Bisswanger, H., (1994) Enzymkinetik, 2nd ed. VCH: Weinheim (ISBN 3527300325); Bergmeyer, H.U., Bergmeyer, J., Graßl, M., eds. (1983-1986) Methods of Enzymatic Analysis, 3rd ed., vol. I-XII, Verlag Chemie: Weinheim; and Ullmann's Encyclopedia of Industrial Chemistry (1987) vol. A9, Enzymes. VCH: Weinheim, p. 352-363.

Example 9: Analysis of the Impact of Combinations of Recombinant Proteins on the Production of a Desired Seed Storage Compound. Seeds from transformed Arabidopsis thaliana plants were analyzed by gas chromatography (GC) for total oil content and fatty acid profile. GC

analysis reveals that Arabidopsis plants transformed with a construct containing a combination of LMPs as described herein.

The results suggest that overexpression of the combination of LMPs as described in Table 9 of Figure 8 allows the manipulation of total seed oil content. As an example, the results of the seed lipid analysis of combinations number 21, 23, 26, 27, 32 & 33 are shown in Figure 6. As controls plants transformed with the empty vector, i.e. pSun2 without the combination of trait genes, were grown together with the plants harbouring the combinations of LMPs and their seeds analysed simultaneously.

As a further example the data shown in table 8 in figure 7 demonstrates that seed oil content of canola seed can significantly be increased by introduction of the combinations of LMPs as listed in table 9 of figure 8. T2 seeds of plants harbouring the combination of LMPs listed in table 8 were analysed for seed oil content by NIRS. Control plants were non-transgenic segregants grown together with the transgenic plants carrying the combination of LMPs.Only lines with an increase of more than 5 % are shown. The p-values shown were calculated using simple t-test.

The effect of the genetic modification in plants on a desired seed storage compound (such as a sugar, lipid or fatty acid) can be assessed by growing the modified plant under suitable conditions and analyzing the seeds or any other plant organ for increased production of the desired product (i.e., a lipid or a fatty acid). Such analysis techniques are well known to one skilled in the art, and include spectroscopy, thin layer chromatography, staining methods of various kinds, enzymatic and microbiological methods, and analytical chromatography such as high performance liquid chromatography (see, for example, Ullman 1985, Encyclopedia of Industrial Chemistry, vol. A2, pp. 89-90 and 443-613, VCH: Weinheim; Fallon, A. et al. 1987, Applications of HPLC in Biochemistry in: Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17; Rehm et al., 1993 Product recovery and purification, Biotechnology, vol. 3, Chapter III, pp. 469-714, VCH: Weinheim; Belter, P.A. et al., 1988 Bioseparations: downstream processing for biotechnology, John Wiley & Sons; Kennedy J.F. & Cabral J.M.S. 1992, Recovery processes for biological materials, John Wiley and Sons; Shaeiwitz J.A. & Henry J.D. 1988, Biochemical separations in: Ulmann's Encyclopedia of Industrial Chemistry, Separation and purification techniques in biotechnology, vol. B3, Chapter 11, pp. 1-27, VCH: Weinheim; and Dechow F.J. 1989).

Besides the above-mentioned methods, plant lipids are extracted from plant material as described by Cahoon et al. (1999, Proc. Natl. Acad. Sci. USA 96, 22:12935-12940) and Browse et al. (1986, Anal. Biochemistry 442:141-145). Qualitative and quantitative lipid or fatty acid analysis is described in Christie, William W., Advances in Lipid Methodology. Ayr/Scotland:Oily Press. - (Oily Press Lipid Library; Christie, William W., Gas Chromatography and Lipids. A

Practical Guide - Ayr, Scotland:Oily Press, 1989 Repr. 1992. - IX,307 S. - (Oily Press Lipid Library; and "Progress in Lipid Research," Oxford :Pergamon Press, 1 (1952) – 16 (1977) Progress in the Chemistry of Fats and Other Lipids CODEN.

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

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 2nd Edition, Pergamon Press, Exeter, ISBN 0-08-023791-6).

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, 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 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 14C-acetate or 14C-pyruvate (see, e.g. Focks & Benning 1998, Plant Physiol. 118:91-101; Eccleston & Ohlrogge 1998, Plant Cell 10:613-621). The distribution of carbon-14 into lipids and aqueous soluble components can be determined by liquid scintillation counting after the respective separation (for example on TLC plates) including standards like 14C-sucrose and 14C-malate (Eccleston & Ohlrogge 1998, Plant Cell 10:613-621).

Material to be analyzed can be disintegrated via sonification, glass milling, liquid nitrogen, and grinding, or via other applicable methods. The material has to be centrifuged after disintegration. The sediment is re-suspended in distilled water, heated for 10 minutes at 100°C, cooled on ice and centrifuged again followed by extraction in 0.5 M sulfuric acid in methanol

containing 2% dimethoxypropane for 1 hour at 90°C leading to hydrolyzed oil and lipid compounds resulting in transmethylated lipids. These fatty acid methyl esters are extracted in petrolether and finally subjected to GC analysis using a capillary column (Chrompack, WCOT Fused Silica, CP-Wax-52 CB, 25 m, 0.32 mm) at a temperature gradient between 170°C and 240°C for 20 minutes and 5 min. at 240°C. The identity of resulting fatty acid methylesters is defined by the use of standards available form commercial sources (i.e., Sigma).

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

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).

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

To quantify soluble sugars, 10 µl of the sugar extract is added to 990 µl of reaction buffer containing 100 mM imidazole, pH 6.9, 5 mM MgCl2, 2 mM NADP, 1 mM ATP, and 2 units 2 ml-1 of Glucose-6-P-dehydrogenase. For enzymatic determination of glucose, fructose, and sucrose, 4.5 units of hexokinase, 1 unit of phosphoglucoisomerase, and 2 µl of a saturated fructosidase solution are added in succession. The production of NADPH is photometrically monitored at a wavelength of 340 nm. Similarly, starch is assayed in 30 µl 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 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 µl of acetone in a 1.5-

ml polypropylene test tube. Following centrifugation at 16,000g, the supernatant is discarded and the vacuum-dried pellet is resuspended in 250 µl of extraction buffer containing 50 mM Tris-HCl, pH 8.0, 250 mM NaCl, 1 mM EDTA, and 1% (w/v) SDS. Following incubation for 2 h at 25°C, the homogenate is centrifuged at 16,000g for 5 min and 200 ml of the supernatant will be used for protein measurements. In the assay, γ-globulin is used for calibration. For protein measurements, Lowry DC protein assay (Bio-Rad) or Bradford-assay (Bio-Rad) is used.

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

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).

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).

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

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.

Additionally, the combinations of 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 at al. 1998, Plant J. 15:39-48). The resultant knockout cells can then be evaluated for their composition and content in seed storage compounds, and the effect on the phenotype and/or genotype of the mutation. For other

methods of gene inactivation include US 6004804 “Non-Chimeric Mutational Vectors” and Puttaraju et al. (1999, “Spliceosome-mediated RNA trans-splicing as a tool for gene therapy,” Nature Biotech. 17:246-252).

Example 10: Purification of the Desired Products from Transformed Organisms. LMPs 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), Ulmann’s Encyclopedia of Industrial Chemistry (1996, Vol. A27, VCH: Weinheim, p. 89-90, p. 521-540, p. 540-547, p. 559-566, 575-581 and p. 581-587) and Michal G. (1999, Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley and Sons; Fallon, A. et al. 1987, Applications of HPLC in Biochemistry in: Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17).

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.

Table 2. Plant Lipid Classes

Neutral Lipids	Triacylglycerol (TAG)
	Diacylglycerol (DAG)
	Monoacylglycerol (MAG)
Polar Lipids	Monogalactosyldiacylglycerol (MGDG)
	Digalactosyldiacylglycerol (DGDG)
	Phosphatidylglycerol (PG)
	Phosphatidylcholine (PC)
	Phosphatidylethanolamine (PE)
	Phosphatidylinositol (PI)
	Phosphatidylserine (PS)
	Sulfoquinovosyldiacylglycerol

Table 3. 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
γ-18:3	Gamma-linolenic acid*
20:0	Arachidic acid
20:1	Eicosenoic acid
22:6	Docosahexanoic acid (DHA) *
20:2	Eicosadienoic acid
20:4	Arachidonic acid (AA) *
20:5	Eicosapentaenoic acid (EPA) *
22:1	Erucic acid

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

Table 4. A table of the putative functions of the LMPs (the full length nucleic acid sequences can be found in Appendix A using the sequence codes, column 2 shows the concordance of the sequence identifier used in Appendix A with or the sequence identifier of the WIPO Standard ST. 25 sequence listing)

Seq ID as used in Appendix A	SEQ ID as used in WIPO Standard ST. 25 sequence listing	Sequence name	Species	Function
1	1	Wri	Arabidopsis thaliana	wrinkle transcription factor
2	3	JB05	Arabidopsis thaliana	beta-ketoacyl-CoA synthase
3	5	JB4054	Arabidopsis thaliana	enoyl CoA hydratase/isomerase
4	7	CTR1	Arabidopsis thaliana	Regulator of ethylene response
5	9	CK	Physcomitrella patens	Protein kinase
6	11	DGD	Arabidopsis thaliana	Phospholipid metabolism
7	13	Susy	Arabidopsis thaliana	Sucrose synthase
8	15	PCT	Arabidopsis thaliana	Phospholipid metabolism

Table 5 with concordance of sequence identifiers used for promoters of Appendix A

Seq ID as used in Appendix A	SEQ ID as used in WIPO Standard ST. 25 sequence listing	Sequence name
9	17	PtxA
10	18	USP
11	19	LeB4
12	20	LEB4
13	21	Conlinin

Table 6 with concordance of sequence identifiers used for terminators of Appendix A

Seq ID as used in Appendix A	SEQ ID as used in WIPO Standard ST. 25 sequence listing	Sequence name
14	22	E9
15	23	A7
16	24	OCS
17	25	LeBT

We claim:

1. An isolated nucleic acid comprising two or more LMP polynucleotide sequences selected from the group consisting of:
   a. a polynucleotide sequence as described by SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7 or SEQ ID NO: 8;
   b. a polynucleotide sequence encoding a polypeptide as described by SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20 or SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24 or SEQ ID NO: 25;
   c. a polynucleotide sequence having at least 70% sequence identity with the nucleic acid of a) or b) above;
   d. a polynucleotide sequence that is complementary to the nucleic acid of a) or b) above; and
   e. a polynucleotide sequence that hybridizes under stringent conditions to nucleic acid of a) or b) above.

2. An isolated polypeptide encoded by a polynucleotide sequence as claimed in claim 1.

3. The isolated nucleic acid of claim 1, wherein the isolated nucleic acid encodes a polypeptide that functions as a modulator of a seed storage compound in microorganisms or plants.

4. The isolated polypeptide of claim 2, wherein the isolated polypeptide sequence functions as a modulator of a seed storage compound in microorganisms or plants.

5. An expression vector containing the nucleic acid of Claim 1, wherein the nucleic acid is operatively linked to a promoter selected from the group consisting of a seed-specific promoter, a root-specific promoter, and a non-tissue-specific promoter.

6. The expression vector of claim 5, wherein the promoter is selected from the group consisting of:

   a. a polynucleotide sequence as described by SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12 or SEQ ID NO: 13;

b. a polynucleotide sequence having at least 70% sequence identity with the nucleic acid of a) above;

c. a polynucleotide sequence that hybridizes under stringent conditions to nucleic acid of a) or b) above;

d. a polynucleotide sequence comprising at least 50 % of the polynucleotide sequences as described by the capital letters of the polynucleotide sequence as described by SEQ ID NO: 26-156 related to a promoter as described by columns 1 and 2 of table 10; and

e. a polynucleotide sequence comprising a polynucleotide sequence having at least 70% sequence identity with the full length polynucleotide sequence as described by the capital letters of the polynucleotide sequence as described by SEQ ID NO: 26-156, wherein the polynucleotide sequence comprises 50% of the nucleotide sequences having at least 70% sequence identity with the polynucleotide sequence as described by the capital letters of the polynucleotide sequence as described by SEQ ID NO: 26-156 related to a promoter as described by columns 1 and 2 of table 10.

7. The expression vector of claim 5, wherein the terminator is selected form the group consisting of:

a. a polynucleotide sequence as described by SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16 or SEQ ID NO: 17;

b. a polynucleotide sequence having at least 70% sequence identity with the nucleic acid of a) above;

c. a polynucleotide sequence that is complementary to the nucleic acid of a) or b) above; and

d. a polynucleotide sequence that hybridizes under stringent conditions to nucleic acid of a) or b) above.

8. A method of producing a transgenic plant having a modified level of a seed storage compound weight percentage compared to an empty vector control,

a. a first step of introduction into a plant cell of an expression vector containing a nucleic acid, and

b. a further step of generating from the plant cell the transgenic plant,

wherein the nucleic acid encodes a polypeptide that functions as a modulator of a seed storage compound in the plant, and wherein the nucleic acid comprises a polynucleotide sequence of claim 1.

9. The method of Claim 8, wherein the nucleic acid comprises a polynucleotide sequence having at least 90% sequence identity with the polynucleotide sequence of a) or b) of Claim 1.

10. The method of Claim 8, wherein the total seed oil content weight percentage is increased in the transgenic plant as compared to an empty vector control

11. A method of modulating the level of a seed storage compound weight percentage in a plant comprising, modifying the expression of a nucleic acid in the plant, comprising

a. a first step of introduction into a plant cell of an expression vector comprising a nucleic acid, and
b. a further step of generating from the plant cell the transgenic plant,

wherein the nucleic acid encodes a polypeptide that functions as a modulator of a seed storage compound in the plant wherein the nucleic acid comprises the polynucleotide sequence of claim 1.

12. The method of Claim 11, wherein the total seed oil content weight percentage is increased in the transgenic plant as compared to an empty vector control.

13. A transgenic plant made by a method of claims 8 or 11.

14. The transgenic plant of Claim 13, wherein the total seed oil content weight percentage is increased in the transgenic plant as compared to an empty vector control.

15. The transgenic plant of Claim 13, wherein the plant is selected from the group consisting of rapeseed, canola, linseed, soybean, sunflower, maize, oat, rye, barley, wheat, rice, pepper, tagetes, cotton, oil palm, coconut palm, flax, castor, sugarbeet, rice and peanut.

16. A seed produced by the transgenic plant of Claim 13, wherein the plant expresses the polypeptide that functions as a modulator of a seed storage compound and wherein the

plant is true breeding for a modified level of seed storage compound weight percentage as compared to an empty vector control.

FIGURE 1:
Open reading frames of the nucleic acid sequences coding for LMP usefull in novel combinations to increase oil content:

1A: Wri (SEQ ID NO. 1):
atgaagaagcgcttaaccacttccacttgttcttcttctccatcttcctctgtttcttcttctactactacttcctctcctattcagtcggagg
ctccaaggcctaaacgagccaaaagggctaagaaatcttctccttctggtgataaatctcataacccgacaagccctgcttcta
cccgacgcagctctatctacagaggagtcactagacatagatggactgggagattcgaggctcatctttgggacaaaagctctt
ggaattcgattcagaacaagaaaggcaaacaagtttatctgggagcatatgacagtgaagaagcagcagcacatacgtacg
atctggctgctctcaagtactggggacccgacaccatcttgaattttccggcagagacgtacacaaaggaattggaagaaatg
cagagagtgacaaaggaagaatatttggcttctctccgccgccagagcagtggtttctccagaggcgtctctaaatatcgcggc
gtcgctaggcatcaccacaacggaagatgggaggctcggatcggaagagtgtttgggaacaagtacttgtacctcggcacct
ataatacgcaggaggaagctgctgcagcatatgacatggctgcgattgagtatcgaggcgcaaacgcggttactaatttcgac
attagtaattacattgaccggttaaagaagaaaggtgttttcccgttccctgtgaaccaagctaaccatcaagagggtattcttgtt
gaagccaaacaagaagttgaaacgagagaagcgaaggaagagcctagagaagaagtgaaacaacagtacgtggaag
aaccaccgcaagaagaagaagagaaggaagaagagaaagcagagcaacaagaagcagagattgtaggatattcaga
agaagcagcagtggtcaattgctgcatagactcttcaaccataatggaaatggatcgttgtggggacaacaatgagctggcttg
gaacttctgtatgatggatacagggttttctccgtttttgactgatcagaatctcgcgaatgagaatcccatagagtatccggagcta
ttcaatgagttagcatttgaggacaacatcgacttcatgttcgatgatgggaagcacgagtgcttgaacttggaaaatctggattgt
tgcgtggtgggaagagagagcccaccctcttcttcttcaccattgtcttgcttatctactgactctgcttcatcaacaacaacaaca
acaacctcggtttcttgtaactatttggtctga

1B JB05 (SEQ ID NO. 2):
atggacggtgccggagaatcacgactcggtggtgatggtggtggtgatggttctgttggagttcagatccgacaaacacggatg
ctaccggattttctccagagcgtgaatctcaagtatgtgaaattaggttaccattacttaatctcaaatctcttgactctctgtttattccc
tctcgccgttgttatctccgtcgaagcctctcagatgaacccagatgatctcaaacagctctggatccatctacaatacaatctggt
tagtatcatcatctgttcagcgattctagtcttcgggttaacggtttatgttatgacccgacctagacccgtttacttggttgatttctcttg
ttatctcccacctgatcacctcaaagctccttacgctcggttcatggaacattctagactcaccggagatttcgatgactctgctctc
gagtttcaacgcaagatccttgagcgttctggtttaggggaagacacttatgtccctgaagctatgcattatgttccaccgagaattt
caatggctgctgctagagaagaagctgaacaagtcatgtttggtgctttagataacctttcgctaacactaatgtgaaaccaaa
ggatattggaatccttgttgtgaattgtagtctctttaatccaactccttcgttatctgcaatgattgtgaacaagtataagcttagaggt
aacattagaagctacaatctaggcggtatgggttgcagcgcgggagttatcgctgtggatcttgctaaagacatgttgttggtaca
taggaacacttatgcggttgttgtttctactgagaacattactcagaattggtattttggtaacaagaaatcgatgttgataccgaact
gcttgtttcgagttggtggctctgcggttttgctatcgaacaagtcgagggacaagagacggtctaagtacaggcttgtacatgta
gtcaggactcaccgtggagcagatgataaagctttccgttgtgtttatcaagagcaggatgatacagggagaaccggggtttcg
ttgtcgaaagatctaatggcgattgcaggggaaactctcaaaaccaatatcactacattgggtcctcttgttctaccgataagtga
gcagattcccttctttatgactctagttgtgaagaagctctttaacggtaaagtgaaaccgtatatcccggatttcaaacttgctttcg
agcatttctgtatccatgctggtggaagagctgtgatcgatgagttagagaagaatctgcagctttcaccagttcatgtcgaggctt
cgaggatgactcttcatcgatttggtaacacatcttcgagctccatttggtatgaattggcttacattgaagcgaagggaaggatg
cgaagaggtaatcgtgtttggcaaatcgcgttcggaagtggatttaaatgtaatagcgcgatttgggaagcattaaggcatgtga
aaccttcgaacaacagtccttgggaagattgtattgacaagtatccggtaactttaagttattag

1C JB4054 (SEQ ID NO. 3):
atgtgttcattagagaaacgtgatcgtcttttcatactaaaactcaccggcgacggcgaacaccgtctaaacccaaccttattcga
ctctctccgctccaccatcaaccaaatccgatcagatccatcattttcacaatcagtactcatcacaacatcagatggtaaattctt
ctccaacggctacgatctcgctttagccgagtcaaatccttctctctctgttgtaatggacgcaaaacttagatccttagtcgccgat
ctaatctctcttcctatgccaacaatcgccgccgtcacaggtcacgcttccgccgcgggatgtatttagcgatgagtcatgattat

gtattgatgcgtcgtgatagaggttttttgtatatgagtgaattggatattgagttgatagttccggcgtggttcatggctgttattaggg
gtaagattggttctccggcggccagaagggatgtgatgttgacggcggcgaaagtgacggcggatgtgggtgttaagatggg
gattgttgattcggcgtatggtagtgcggcggagacggttgaagccgccattaagttagatgaggagattgttcagagaggtggt
gatggacacgtgtatggtaagatgagagagagtcttttaagagaggttcttattcatacgattggtgaatatgagagtggttcaagt
gtggtgcgtagcactggatctaaacttag

1D CTR1 (SEQ ID NO. 4):
atggaaatgcccggtagaagatctaattacactttgcttagtcaattttctgacgatcaggtgtcagtttccgtcaccggagctcctc
cgcctcactatgattccttgtcgagcgaaaacaggagcaaccataacagcgggaacaccgggaaagctaaggcggagag
aggcggatttgattgggatcctagcggtggtggtggtggtgatcataggttgaataatcaaccgaatcgggttgggaataatatgt
atgcttcgtctctagggttgcaaaggcaatccagtgggagtagtttcggtgagagctctttgtctggggattattacatgcctacgctt
tctgcggcggctaacgagatcgaatctgttggatttcctcaagatgatgggtttaggcttggatttggtggtggtggaggagatttg
aggatacagatggcggcggactccgctggagggtcttcatctgggaagagctgggcgcagcagacggaggagagttatca
gctgcagcttgcattggcgttaaggctttcgtcggaggctacttgtgccgacgatccgaactttctggatcctgtaccggacgagtc
tgctttacggacttcgccaagttcagccgaaaccgtttcacatcgtttctgggttaatggctgcttatcgtactatgataaagttcctg
atgggttttatatgatgaatggtctggatccctatatttggaccttatgcatcgacctgcatgaaagtggtcgcatcccttcaattgaat
cattaagagctgttgattctggtgttgattcttcgcttgaagcgatcatagttgataggcgtagtgatccagccttcaaggaacttcac
aatagagtccacgacatatcttgtagctgcattaccacaaaagaggttgttgatcagctggcaaagcttatctgcaatcgtatggg
gggtccagttatcatgggggaagatgagttggttcccatgtggaaggagtgcattgatggtctaaaagaaatctttaaagtggtg
gttcccataggtagcctctctgttggactctgcagacatcgagctttactcttcaaagtactggctgacataattgatttaccctgtcg
aattgccaaaggatgtaaatattgtaatagagacgatgccgcttcgtgccttgtcaggtttgggcttgatagggagtacctggttga
tttagtaggaaagccaggtcacttatgggagcctgattccttgctaaatggtccttcatctatctcaatttcttctcctctgcggtttcca
cgaccaaagccagttgaacccgcagtcgattttaggttactagccaaacaatatttctccgatagccagtctcttaatcttgttttcg
atcctgcatcagatgatatgggattctcaatgtttcataggcaatatgataatccgggtggagagaatgacgcattggcagaaaa
tggtggtgggtctttgccacccagtgctaatatgcctccacagaacatgatgcgtgcgtcaaatcaaattgaagcagcacctatg
aatgccccaccaatcagtcagccagttccaaacagggcaaatagggaacttggacttgatggtgatgatatggacatccgtg
gtgtgatcttaatataaaagaaaagattggagcaggttcctttggcactgtccaccgtgctgagtggcatggctcggatgttgctgt
gaaaattctcatggagcaagacttccatgctgagcgtgttaatgagttcttaagagaggttgcgataatgaaacgccttcgccac
cctaacattgttctcttcatgggtgcggtcactcaacctccaaatttgtcaatagtgacagaatatttgtcaagaggtagtttatacag
acttttgcataaaagtggagcaagggagcaattagatgagagacgtcgccttagtatggcttatgatgtggctaagggaatgaa
ttatcttcacaatcgcaatcctccaattgtgcatagagatctaaaatctccaaacttattggttgacaaaaaatatacagtcaaggtt
tgtgatttttggtctctcgcgattgaaggccagcacgtttctttcctcgaagtcagcagctggaacccccgagtggatggcaccaga
agtcctgcgagatgagccgtctaatgaaaagtcagatgtgtacagcttcggggtcatcttgtgggagcttgctacattgcaacaa
ccatgggtaacttaaatccggctcaggttgtagctgcggttggtttcaagtgtaaacggctggagatcccgcgtaatctgaatcc
tcaggttgcagccataatcgagggttgttggaccaatgagccatggaagcgtccatcatttgcaactataatggacttgctaaga
ccattgatcaaatcagcggttcctccgcccaaccgctcggatttgtaa

1E CK(SEQ ID NO. 5):
atggaaccccgcgtcggcaacaagtatcgccttggccggaaaattgggagtggttcctttggtgagatctacctggggaccaat
ctcgtgactcatgaggaggtcggcatcaagctggagagcatcaaggccaagcatccacaattgctttatgagtccaagttgtac
cgtattcttcaaggaggaactgggattcccaacatcagatggtacggaattgaaggagactataatgtgatggttcttgatcttctg
ggacccagtcttgaagatcttttcaatttctgcagccggaaattctctttgaagacagttctcatgcttgccgaccagctgatcaatc
gagtggagtatgtgcatgccaagagtttcctccacagggacataaagcctgacaatttcttgatggggctaggcaggcgagca
aatcaggtctatatgattgactttggtcttgcaaagaagtatcgcgatcccactactcatcagcacattccttatagagagaacaa
aaatcttactggaaccgctcgatatgcaagtatcaacactcatcttggtattgaacaaagcaggagagatgatctggagtctcttg
gatatgttctcatgtatttcttgagaggcagcctgccttggcaaggaatgaaagcaggaaccaagaagcagaagtatgaaaaa

atcagtgagaaaaagatgtccacccctatagagttcctttgtaaagcttacccgtctgagtttgcttcatacttccactactgtcggtc
tcttcggttcgatgacaaaccggactatgcttacctgaagagaattttccgagatctcttcattcgtgagggttttcagtttgattatgttt
tcgactggacgattttgaagtatcagcaaacacatttttctggtggtcctctccgtccagcggctgcggcgggaggttcaagtgga
gcagcagcagcagcggcagcaggaattggtacagtcccaagagacgcccagcgagcaattgagcctactgatgttgccgct
cgaactcgaatggttggtgcgactcgctctagtggattaaatccactggacgcgtcaaagcataagagtactagcccagatga
agccgcttctaaggacatagcccttagcggtcttgcagaaccagagcgcacgcatgcttcttcgtttgtgcgggggagctcatca
tcaaggagagctgttgttggatgtgctaggccagcagggtcaacagaggcgggagatggaacgcgggtgttggctggcaaa
atgggccccactagcctgcgcacatcagcaggaatgcagaggagctctccggtggcatctacggatcccaagcggacggg
acgagattcttatgctggaaactccggaagaaatcctagttcctctcgaaattcgaaagagtga

1F DGD (SEQ ID NO. 6):
atggtaaaggaaactctaattcctccgtcatctacgtcaatgacgaccggaacatcttcttcttcgtctctttcaatgacgttatcctc
aacaaacgcgttatcgtttttgtcgaaaggatggagagaggtatgggattcagcagatgcggatttgcagctgatgcgagacag
agctaactctgttaagaatctagcatcaacgttcgatagagagatcgagaatttcctcaataactcggcgaggtctgcgtttcccg
ttggttcaccatcggcgtcgtctttctcaaatgaaattggtatcatgaagaagcttcagccgaagatttcggagtttcgtagggtttat
tcggcgccggagattagtcgcaaggttatggagagatggggacctgcgagagcgaagcttggaatggatctatcggcgatta
agaaggcgattgtgtctgagatggaattggatgagcgtcagggagttttggagatgagtagattgaggagacggcgtaatagt
gatagggttaggtttacggagtttttcgcggaggctgagagagatggagaagcttatttcggtgattgggaaccgattaggtctttg
aagagtagatttaaagagtttgagaaacgaagctcgttagaaatattgagtggattcaagaacagtgaatttgttgagaagctca
aaaccagctttaaatcaatttacaaagaaactgatgaggctaaggatgtccctccgttggatgtacctgaactgttggcatgtttg
gttagacaatctgaaccttttcttgatcagattggtgttagaaaggatacatgtgaccgaatagtagaaagcctttgcaaatgcaa
gagccaacaactttggcgtctgccatctgcacaagcatccgatttaattgaaaatgataaccatggagttgatttggatatgagg
atagccagtgttcttcaaagcacaggacaccattatgatggtgggttttggactgattttgtgaagcctgagacaccggaaaaca
aaaggcatgtggcaattgttacaacagctagtcttccttggatgaccggaacagctgtaaatccgctattcagagcggcgtatttg
gcaaaagctgcaaaacagagtgttactctcgtggttccttggctctgcgaatctgatcaagaactagtgtatccaaacaatctca
ccttcagctcacctgaagaacaagagagttatatacgtaaatggttggaggaaaggattggtttcaaggctgattttaaaatctcc
ttttacccaggaaagttttcaaaagaaaggcgcagcatatttcctgctggtgacacttctcaatttatatcgtcaaaagatgctgac
attgctatacttgaagaacctgaacatctcaactggtattatcacggcaagcgttggactgataaattcaaccatgttgttggaatt
gtccacacaaactacttagagtacatcaagagggagaagaatggagctcttcaagcatttttgtgaaccatgtaaacaattgg
gtcacacgagcgtattgtgacaaggttcttcgcctctctgcggcaacacaagatttaccaaagtctgttgtatgcaatgtccatggt
gtcaatcccaagttccttatgattggggagaaaattgctgaagagagatcccgtggtgaacaagctttctcaaaaggtgcatact
tcttaggaaaaatggtgtgggctaaaggatacagagaactaatagatctgatggctaaacacaaaagcgaacttgggagcttc
aatctagatgtatatgggaacggtgaagatgcagtcgaggtccaacgtgcagcaaagaaacatgacttgaatctcaatttcctc
aaaggaagggaccacgctgacgatgctcttcacaagtacaaagtgttcataaaccccagcatcagcgatgttctatgcacagc
aaccgcagaagcactagccatggggaagtttgtggtgtgtgcagatcacccttcaaacgaattctttagatcattcccgaactgc
ttaacttacaaaacatccgaagactttgtgtccaaagtgcaagaagcaatgacgaaagagccactacctctcactcctgaaca
aatgtacaatctctcttgggaagcagcaacacagaggttcatggagtattcagatctcgataagatcttaaacaatggagaggg
aggaaggaagatgcgaaaatcaagatcggttccgagctttaacgaggtggtcgatggaggattggcattctcacactatgttct
aacagggaacgatttcttgagactatgcactggagcaacaccaagaacaaaagactatgataatcaacattgcaaggatctg
aatctcgtaccacctcacgttcacaagccaatcttcggctggtag

1G Susy (SEQ ID NO. 7):
atgtgtgttgtgattggtctcaagtcatgggtaatggttttggttgttatctttattagatatgtagcccagggaaaggggatattgcagt
cccaccagctgattgatgagttccttaagactgtgaaagttgatggaacattagaagatcttaacaaaagtccattcatgaaagtt
ctgcagtctgcagaggaagccatagttttgcctccatttgttgctttggctatacgtcccagacctggtgttagggaatatgtccgtgt
gaatgtgtatgagctgagcgtagatcatttaactgtttctgaatatcttcggtttaaggaagagctcgttaatggccatgccaatgga

gattatctccttgaacttgattttgaacctttcaatgcaacattgcctcgcccaactcgttcatcatccattgggaatggggttcagttc
ctcaatcgtcacctctcttcaattatgttccgtaacaaagaaagcatggagcctttgcttgagtttctccgcactcacaaacatgatg
gccgtcctatgatgctgaatgatcgaatacagaatatccccatacttcagggagctttggcaagagcagaggagttcctttctaa
acttcctctggcaacaccatactctgaattcgaatttgaactacaagggatgggatttgaaaggggatggggtgacacagcaca
gaaggtttcagaaatggtgcatcttcttctggacatactccaggcacctgatccttctgtcttggagacgtttctaggaaggattcct
atggtgttcaatgttgtgattttgtctccgcatggttactttggccaagccaatgtcttgggtctgcctgatactggtggacaggttgtct
acattcttgatcaagtacgtgcattggaaaatgagatgctccttaggatacagaagcaaggactggaagttattccaaagattct
cattgtaacaagactgctacccgaagcaaagggaacaacgtgcaaccagaggttagaaagagttagtggtacagaacacg
cacacattctgcgaataccatttaggactgaaaagggaattcttcgcaagtggatctcaaggtttgatgtctggccatacctggag
acttttgcagaggatgcatcaaatgaaatttctgcggagttgcagggtgtaccaaatctcatcattggcaactacagtgatggaa
atctcgttgcttctttgttagctagtaagctaggtgtgatacagtgtaatattgctcatgctttagagaaaaccaagtaccccgagtct
gacatttactggagaaaccatgaagataagtatcacttttcaagtcagttcactgcagatctaattgccatgaataatgccgatttc
atcatcaccagcacataccaagagattgcgggaagcaagaacaatgttgggcaatacgagagccacacagctttcactatg
cctggtctttaccgagttgttcatggaattgatgtctttgatcctaagtttaatatggtctctccaggagctgatatgaccatatactttcc
atattccgacaaggaaagaagactcactgcccttcatgagtcaattgaagaactcctctttagtgccgaacagaatgatgagca
tgttggtttactgagcgaccaatcgaagccaatcatcttctctatggcaagacttgacagggtgaaaaacttgactgggctagttg
aatgctatgccaagaatagcaagcttagagagcttgcaaatcttgttatagtcggtggctacatcgatgagaatcagtccaggg
atagagaggaaatggctgagatacaaaagatgcacagcctgattgagcagtatgatttacacggtgagtttaggtggatagct
gctcaaatgaaccgtgctcgaaatggtgagctttaccgttatatcgcagacacaaaaggtgtttttgttcagcctgctttctatgaag
catttgggcttacggttgtggaatcaatgacttgtgcactcccaacgtttgctacctgtcatggtggacccgcagagattatcgaaa
acggagtttctgggttccacattgacccatatcatccagaccaggttgcagctaccttggtcagcttctttgagacctgtaacacca
atccaaatcattgggttaaaatctctgaaggagggctcaagcgaatctatgaaaggtacacatggaagaagtactcagagag
actgcttaccctggctggagtctatgcattctggaaacatgtgtctaagctcgaaaggagagaaacacgacgttacctagagat
gttttactcattgaaatttcgtgatttggccaattcaatcccgctggcaacagatgagaactga

1H PCT (SEQ ID NO. 8):
atggcgacttttgctgaacttgttttatcgacttctcgctgtacatgcccttgccgttcattcactagaaaaccccctaattcgtcccccttt
atctggtctgcgtctccccggtgataccaaaccattgtttcgttccggacttggtcggatttctgttagccggcgtttcctcacggccgt
tgctcgagctgaatcagaccagcttggtgatgatgaccactcaaagggaattgatagaatccataacttgcagaatgtggaag
ataagcagaagaaagcaagccagcttaagaaaagagtgatctttggtattggcattggtttacctgttggatgtgttgtgttagctg
gaggatgggttttcactgtagctttagcatcttctgtttttatcggttcccgcgaatatttcgagcttgttagaagtagaggcatagcta
aaggaatgactcctcctccacgatatgtatctcgagtttgctcggttatatgtgcccttatgcccatacttacactgtactttggtaaca
ttgatatattggtgacatctgcagcatttgttgttgcaatagcattgttagtacaaagaggatccccacgttttgctcagctgagtagt
acaatgtttggtctgttttactgtggttatctcccttctttctgggttaagcttcgctgtggtttagctgctcctgcgcttaacactggtatcg
gaaggacatggccaattcttcttggtggtcaagctcattggacagttggacttgtggcaacattgatttctttcagcggtgtaattgc
gacagacacatttgcttttctcggtggaaagacttttggtaggacacctcttactagtattagtcccaagaagacatgggaagga
actattgtaggacttgttggttgtatagccattaccatattactctctaaatatctcagttggccacaatctctgttcagctcagtagcttt
tgggtttcttaacttctttgggtcagtctttggtgatcttactgaatcaatgatcaagcgtgatgctggcgtcaaagactctggttcactt
atcccaggacacggtggaatattagatagagttgatagttacattttcaccggcgcattagcttattcattcatcaaaacatcccta
aaactttacggagtttgatga

FIGURE 2
Nucleic acid sequences of exemplary promoters usefull in novel combinations to increase oil content:

2A PtxA (SEQ ID NO. 9):
Cgcaatttttgtgaagctgagggaggattggattttacacctattcaaaagtcattcaaagtttgtccctccattcaaggatgaatg
tagatttttcaagcatcaaacacaagaatcactagcataacatgctttgaaacccacacacttaaattaatgttaggaatatcaaa
tccaatataaaatcatagttgtcaattacatactcaatcaagtccctttctttttacccaataaacatcaacatattgcttcttccattaag
catataaacatcaaagtctaaaactagcaaaatgttgtttttaggatgacacatttcatacatagtttaaaagatacttgattcgatta
caaaaagaaattaccaatagtttagcacaaagtctaaagcataattaaagcatcacatgtgcagatttatgaaaaaaagatta
agattgccccttttcatcacgggtcgaataatagcactacttgtcactacatgttaaaaaaatgtcctctagtacatcaaacttttttcca
ttgattcccctttatccatgaaaaaaataaacaaattcttaagacacaaaaaaatggccccacatccttttttctggcctagtttgtttg
aattcattctaactcttgaatatgtaacgaggcccactaaaaatcaatcaatgatttaacataaaaaatgaatagtttaattccaatt
tgctgcaacatggtccgtgaatatgactcacgagaaagatatatcaaaatatcaaaatttcatagtttttttcaccatataaacctca
tcactcattctatttttttaagtgcaaagcttcatagtta

2B USP (SEQ ID NO. 10):
caaatttacacattgccactaaacgtctaaaccccttgtaatttgtttttgttttactatgtgtgttatgtatttgatttgcgataaatttttatatt
tggtactaaatttataacacctttttatgctaacgtttgccaacacttagcaatttgcaagttgattaattgattctaaattatttttgtcttct
aaatacatatactaatcaactggaaatgtaaatatttgctaatatttctactataggagaattaaagtgagtgaatatggtaccaca
aggtttggagatttaattgttgcaatgatgcatggatggcatatacaccaaacattcaataattcttgaggataataatggtaccac
acaagatttgaggtgcatgaacgtcacgtggacaaaaggtttagtaatttttcaagacaacaatgttaccacacacaagttttga
ggtgcatgcatggatgccctgtggaaagtttaaaaatattttggaaatgatttgcatggaagccatgtgtaaaaccatgacatcca
cttggaggatgcaataatgaagaaaactacaaatttacatgcaactagttatgcatgtagtctatataatgaggattttgcaatact
ttcattcatacacactcactaagttttacacgattataatttcttcatagc

2C 0.7LeB4 (SEQ ID NO. 11):
gagttaccatttctttttcctgcatctcaatagtatatagggtatcaaatagtgattatccaaacttaaataagttagaggaaacacc
aagatatgccatatactctcatatttgacactatgattcaaagttgcacttgcataaaacttattaattcaatagtaaaaccaaactt
gtgcgtgatacagttaaaatgactaaactactaattaaggtccctcccattagtaaataagttatttttcttagaaaaagaaaataat
aaaaagaatgacgagtctatctaaatcatattaacaagtaatacatattgattcattcgatggaggaggccaataattgtagtaa
acaagcagtgccgaggttaatatatgctcaagacagtaaataatctaaatgaattaagacagtgatttgcaaagagtagatgc
agagaagagaactaaagatttgctgctacacgtatataagaatagcaacagatattcattctgtctctttgtggaatatggatatct
actaatcatcatctatctgtgaagaataaaagaagcggccacaagcgcagcgtcgcacatatgatgtgtatcaaattaggactc
catagccatgcatgctgaagaatgtcacacacgttctgtcacacgtgttactctctcactgttctcctcttcctataaatcaccgcgc
cacagcttctccacttcaccacttcaccacttcactcacaatccttcattagttgtttactatcacagtcacag

2D 2.4LEB4 (SEQ ID NO. 12):
gtggaattcgaggggatctgtcgtctcaaactcattcatcagaaccttcttgaacttagttatctcttgttcagagcttcctgttagca
atatgtcatcaacatataaacatgtcccagaagccagaagatagaagttggatgatagaagtaaagtaatgttactggtggagt
accacaatacaagttcatacaaactttattgtccagaaactaacaaagttgagttcagcatagatgaaagacaaaaagaatat
attaaatgacggctgcaaaataaggagtaatgaatacattgacctacctactactaggctatttatacacaatattagggtataat
aaaatattaaaataccctctatcagacttagtcaataagacattcctaaaatataaattatttccaacaataatttgtctcaaataaa
atatagaggtgcaaaagttaaactaagagtgcaaagtaaaattttgagagggctcaaaattgaatataataacaatattagtgt
agtttaagaaaactcaggggatgcagttgaactccctcaactgtacgtagctcctcccctggatgcagtgtaaagatttgaagat
atattttagtactttggatattgtaggccagagggtgttgaagataaaggttcaggaactaacacattcatccacaacttctatgtgt

ccatcgtcagtgaaatacatgccaaataggggagttaagaagagtagaaagggtcaagatagtgatgtgcatcgtgatccttc
ataatgggagtgtggtgagggctcgcatgggagtcatactacaaagagatcatgcataaaaccaactagaagtcaactgtca
agtatgacggctgacaattaaccgtccaccaaatcttccagacatgtttacttgtcccagttttctgatttcttatatccatacattgatg
acattattgatgttggtggcgatggagattggggttttcatgctattacagctttacttggatggggtgaagagtcatagcctttgattc
agacgcagttagatactcaagttcatcaacaccctcaattgtttttaagttgttttgtgacacgatctctacagttagaaatgcgttac
gagtagaacacttggctgtgcagggtatagataaatgaatgacgatttatgatatgggttaccctattgcttctagatacaatgtcgt
atttgtctcccttccaaaagacttaacatcacgtttttcctcttgccttatctccacctatgtatacaagcaggcataaaatcattgttgt
tggttttgtcaacaacaatcattgagtttaggtaaagttgaaacttgattgtccattacctcttgtcactgactgttgaagacagaattg
tactgactgtatatatcaacatatgcgagacgcgttaggcagtggaaagacgtagttaggatgtcatcataatttgtttcgtatttttat
atgtagcacagttttatatgtatatattttatcgggtagttttttatcgattcagttatttgagaaaaagtaatgcagacaaaaagtgga
aaagacaatctgactgtacataagaaatttccaatttttgaaatttttttataattatcagaaattttaaaatttccgataaaaacatac
atgtatagatcgaaaatttcaaatttctagtactttcaaatttcttgcagtaaaagttgtaatttttaaaaatttacgataatttacagtat
ttaaaaaaaaatccaatcttaaataaagggtataagaataaaagcactcatgtggagtggcaggtttcgtcacaccctaagaa
catccctaaatacaccacatatgtataagtattaagtgattgatgttaagtgaaacgaaaatatttatatgtgaaatttaatattcag
cttacttgattaaactccatagtgacccaataagtgctaacttttactgtctttacctttaaatgttatattgatttatttatgcatttctttttcc
tgcatctcaatagtatatagggtatcaaatagtgattatccaaacttaaataagttagaggaaacaccaagatatgccatatactc
tcaaatttgacactatgattcaaagttgcacttgcataaaacttattaattcaatagtaaaaccaaacttgtgcgtgatacagttaaa
atgactaaactactaattaaggtccctcccattagtaaataagttatttttttagaaaaagaaaataataaaaagaatgacgagtc
tatctaaatcatattaacaagtaatacatattgattcattcgatggaggaggccaataattgtagtaaacaagcagtgccgaggtt
aatatatgctcaagacagtaaataatctaaatgaattaagacagtgatttgcaaagagtagatgcagagaagagaactaaag
atttgctgctacacgtatataagaatagcaacagatattcattctgtctctttgtggaatatggatatctactaatcatcatctatctgtg
aagaataaaagaagcggccacaagcgcagcgtcgcacatatgatgtgtatcaaattaggactccatagccatgcatgctgaa
gaatgtcacacacgttctgtcacacgtgttactctctcactgttctcctcttcctataaatcaccgcgccacagcttctccacttcacc
acttcaccacttcactcacaatccttcattagttgtttactatcacagtcaca

2E Conlinin (SEQ ID NO. 13):
ttagcagatatttggtgtctaaatgtttattttgtgatatgttcatgtttgaaatggtggtttcgaaaccagggacaacgttgggatctga
tagggtgtcaaagagtattatggattgggacaatttcggtcatgagttgcaaattcaagtatatcgttcgattatgaaaattttcgaa
gaatatcccatttgagagagtctttacctcattaatgtttttagattatgaaattttatcatagttcatcgtagtctttttggtgtaaaggctg
taaaaagaaattgttcacttttgttttcgtttatgtgaaggctgtaaaagattgtaaaagactattttggtgttttggataaaatgatagtt
tttatagattcttttgcttttagaagaaatacatttgaaatttttccatgttgagtataaaataccgaaatcgattgaagatcatagaaa
tattttaactgaaaacaaatttataactgattcaattctctccatttttatacctatttaaccgtaatcgattctaatagatgatcgatttttt
atataatcctaattaaccaacggcatgtattggataattaaccgatcaactctcacccctaatagaatcagtattttccttcgacgtt
aattgatcctacactatgtaggtcatatccatcgttttaatttttggccaccattcaattctgtcttgcctttagggatgtgaatatgaacg
gccaaggtaagagaataaaaataatccaaattaaagcaagagaggccaagtaagataatccaaatgtacacttgtcattgcc
aaaattagtaaaatactcggcatattgtattcccacacattattaaaataccgtatatgtattggctgcatttgcatgaataatactac
gtgtaagcccaaaagaacccacgtgtagcccatgcaaagttaacactcacgaccccattcctcagtctccactatataaaccc
accatccccaatctcaccaaacccaccacacaactcacaactcactctcacacct

FIGURE 3
Nucleic acid sequences of exemplary terminators usefull in novel combinations to increase oil content:

3A E9 (SEQ ID NO. 14):
ggatcctctagctagagctttcgttcgtatcatcggtttcgacaacgttcgtcaagttcaatgcatcagtttcattgcgcacacacca
gaatcctactgagtttgagtattatggcattgggaaaactgtttttcttgtaccatttgttgtgcttgtaatttactgtgtttttattcggttttc
gctatcgaactgtgaaatggaaatggatggagaagagttaatgaatgatatggtccttttgttcattctcaaattaatattatttgttttt
ctcttatttgttgtgtgttgaatttgaaattataagagatatgcaaacattttgttttgagtaaaaatgtgtcaaatcgtggcctctaatga
ccgaagttaatatgaggagtaaaacacttgtagttgtaccattatgcttattcactaggcaacaaatatattttcagacctagaaaa
gctgcaaatgttactgaatacaagtatgtcctcttgtgttttagacatttatgaactttcctttatgtaattttccagaatccttgtcagattc
taatcattgctttataattatagttatactcatggatttgtagttgagtatgaaaatatttttttaatgcattttatgacttgccaattgattgac
aacatgcatcaatcgaccgggtacc

3B A7 (SEQ ID NO. 15):
ctgaattaacgccgaattaattcgggggatctggattttagtactggattttggttttaggaattagaaattttattgatagaagtatttta
caaatacaaatacatactaagggtttcttatatgctcaacacatgagcgaaaccctataggaaccctaattcccttatctgggaa
ctactcacacattattatggagaaactcgagcttgtcga

3C OCS (SEQ ID NO. 16):
ctgctttaatgagatatgcgagacgcctatgatcgcatgatatttgctttcaattctgttgtgcacgttgtaaaaaacctgagcatgtg
tagctcagatccttaccgccggtttcggttcattctaatgaatatatcacccgttactatcgtatttttatgaataatattctccgttcaatt
tactgattgtcc

3D LeBT (SEQ ID NO. 17):
atcctgcaatagaatgttgaggtgaccactttctgtaataaaataattataaaataaatttagaattgctgtagtcaagaacatcag
ttctaaaatattaataaagttatggccttttgacatatgtgtttcgataaaaaaatcaaaataaattgagatttattcgaaatacaatg
aaagtttgcagatatgagatatgtttctacaaaataataacttaaaactcaactatatgctaatgtttttcttggtgtgtttcatagaaa
attgtatccgtttcttagaaaatgctcgtaa

FIGURE 4

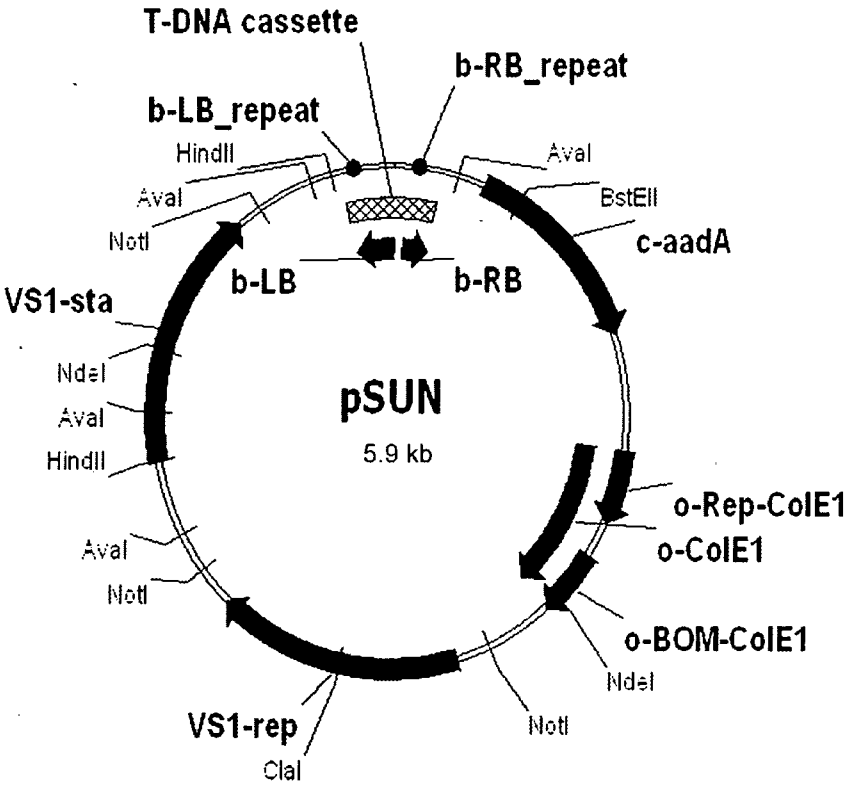


Figure 4: Schematic representation of the binary vector pSUN indicating relevant features and restriction sites. b-RB = right border of T-DNA; c-aadA = aminoglycoside 3'-adenylyl-transferase codons; o-ColE1 replication origin of the plasmid pBR322, consisiting of the two components o-REP-ColE1 and o-BOM-ColE1; VS1-rep = replication origin and repA of plasmid pVS1 VS1-sta = *sta* gene from plasmid pVS1; b-LB = left border of T-DNA;

FIGURE 5

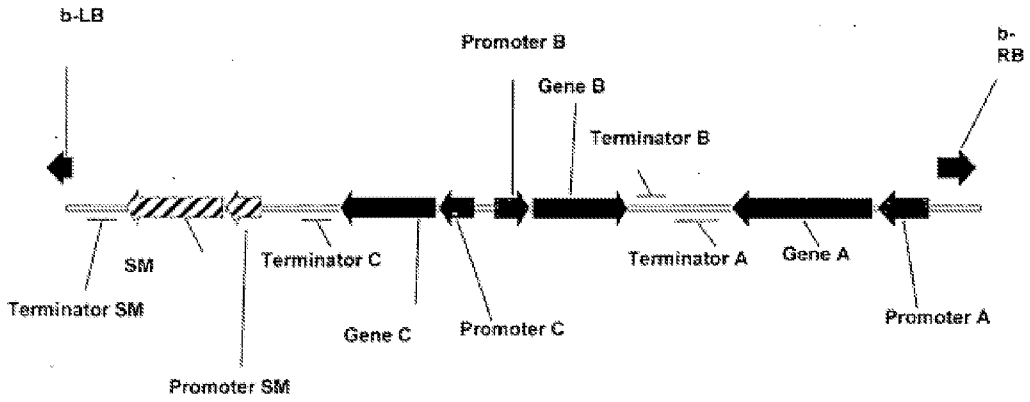
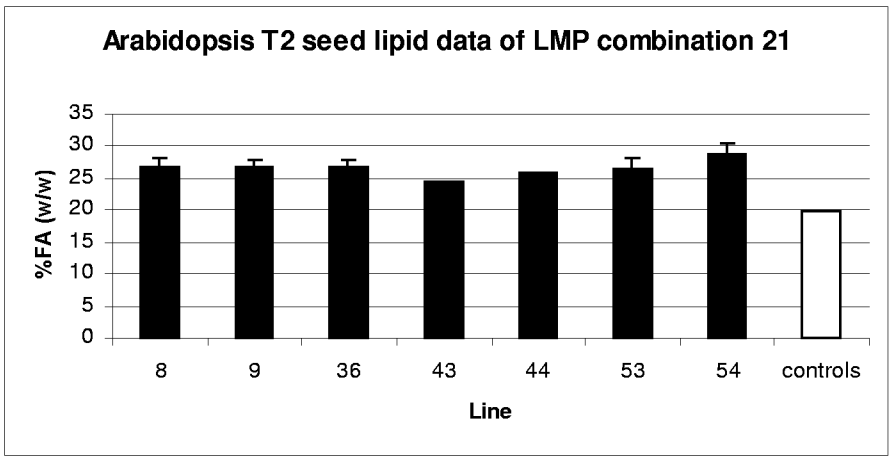
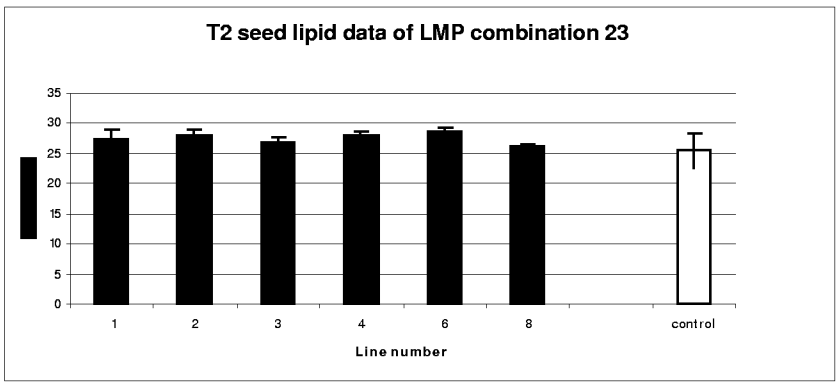

b-LB
Promoter B
Gene B
Terminator B
b-RB
SM
Terminator C
Terminator A
Gene A
Terminator SM
Gene C
Promoter C
Promoter A
Promoter SM

FIGURE 6

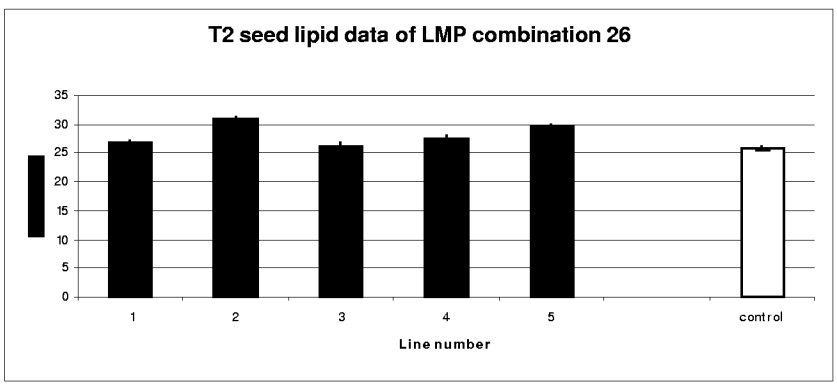
A



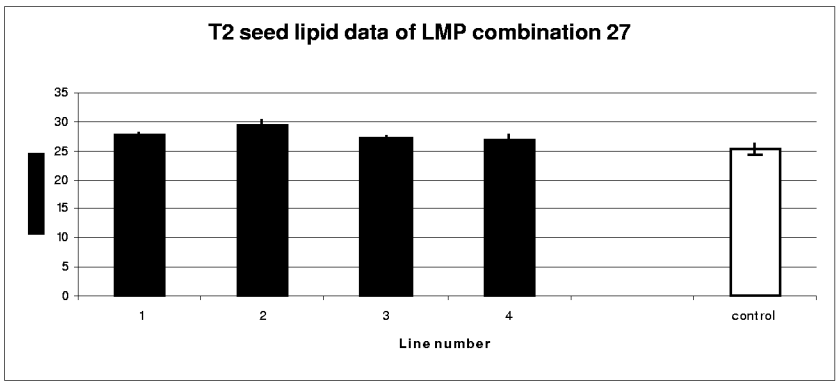
B



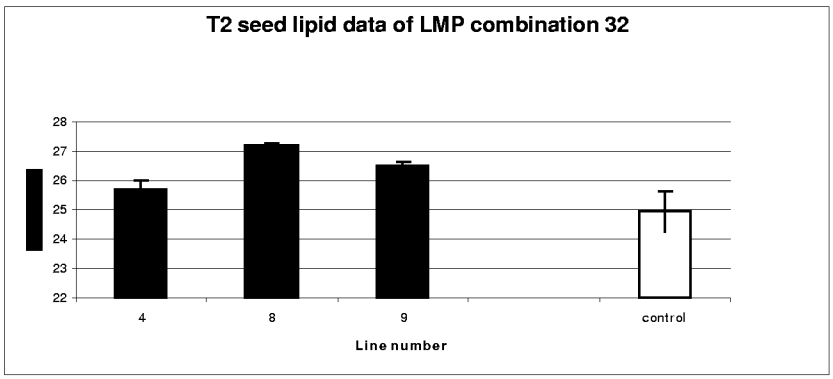
C



D


T2 seed lipid data of LMP combination 27
35
30
25
20
15
10
5
0
1
2
3
4
control
Line number

E


T2 seed lipid data of LMP combination 32
28
27
26
25
24
23
22
4
8
9
control
Line number

F

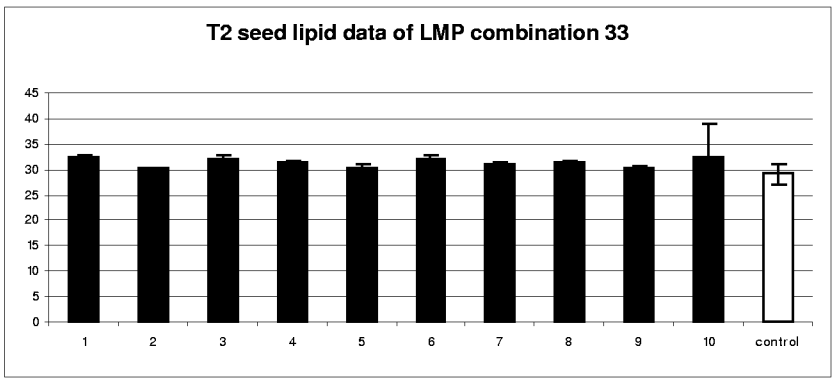

T2 seed lipid data of LMP combination 33
45
40
35
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9
10
control

Table 7: Maximum oil increase observed in T2 Arabidopsis seed of transgenic plants carrying the combinations of LMPs

Combination of LMPs	Maximal relative oil increase observed in a line as % of the control value
23	112,6
26	120,1
27	116,3
32	109,1
33	111,3

Figure 7

Table 8: Increase in canola T2 seed oil of transgenic canola plants carrying the combinations of LMPs

LMP combination	line	relative oil content compared to control	p-Value
42	284	107,3	0,05
23	3	106,0	0,11
27	48	107,0	0,23

Figure 8:

Table 9: A table of combinations of the LMPs as listed in Appendix A

	Position	A		B			C		
Combination number	Promoter	Gene			Gene	Promoter	Promoter	gene	
21	ptxA	WRI	E9	A7	dgd	USP			
22	ptxA	WRI	E9	A7	CK	USP			
23	0.7 LeB4	CTR1	LeB3	A7	dgd	USP			
24	ptxA	WRI	E9	LeB3	PCT	2.7 LeB4			
25	ptxA	WRI	E9	A7	JB05	USP			
26	USP	JB05	A7+OCS	LeB3	CTR1	2.7 LeB4			
27	USP	JB05	A7+OCS	LeB3	CK	2.7 LeB4			
28				LeB3	JB4054	2.7 LeB4	USP	susy	A7
29	ptxA	WRI	E9	LeB3	CK	2.7 LeB4	USP	susy	A7
30	ptxA	WRI	E9	LeB3	CTR1	2.7 LeB4	USP	JB05	A7
31	USP	CK	A7+OCS	LeB3	PCT	2.7 LeB4	Conlinin	susy	E9
32				LeB3	JB4054	2.7 LeB4	USP	JB05	A7
33	ptxA	WRI	E9	LeB3	JB4054	2.7 LeB4	Conlinin	susy	OCS
34	USP	JB05	A7+OCS	LeB3	PCT	2.7 LeB4	Conlinin	susy	E9
35	USP	CK	A7+OCS	LeB3	JB4054	2.7 LeB4	Conlinin	susy	E9
36	USP	CK	A7+OCS	LeB3	CTR1	2.7 LeB4	Conlinin	susy	E9
37	0.7 LeB4	CTR1	LeB3	A7	CK	USP	Conlinin	susy	E9
38	0.7 LeB4	CTR1	LeB3	A7	JB05	USP	Conlinin	susy	E9

39	0.7 LeB4	CTR1	LeB3	A7	dgd	USP	Conlinin	susy	E9
40	USP	CK	A7+OCS	LeB3	DGD	2.7 LeB4	Conlinin	susy	E9
41	USP	JB05	A7+OCS	LeB3	DGD	2.7 LeB4	Conlinin	susy	E9
42	USP	JB05	A7+OCS	LeB3	JB4054	2.7 LeB4			

Figure 9:

Table 10: Transcription Factor Binding Sites of the promoters of the present invention

Promoter_Sequence_ID Sequence identifier of promoter	IUPAC_Sequence_ID Sequenzce identifier of the IUPAC Matrix	TFBS_Sequence_ID Sequence identifier of the Transcription Factor Binding Site	Matrix_ID	Description	Core_Sim (%)	Matrix_Sim (%)
9	44	283	PSATCTA.01	Motif involved in carotenoid and tocopherol biosynthesis and in the expression of photosynthesis-related genes	100,0	92,8
9	45	285	PSATHB1.01	Arabidopsis thaliana homeo box protein 1	100,0	90,7
9	46	303	PSATHB5.01	HDZip class I protein ATHB5	82,9	94,0
9	46	304	PSATHB5.01	HDZip class I protein ATHB5	93,6	93,6
9	48	313	PSATML1.01	L1-specific homeodomain protein ATML1 (A. thaliana meristem layer 1)	75,0	84,0

WO 2007/012576 PCT/EP2006/064276

9	54	301 P$ATTCP20.01	TCP class I transcription factor (Arabidopsis)	100,0	95,6
9	59	260PSCAAT.01	CCAAT-box in plant promoters	100,0	98,5
9	59	260PSCAAT.01	CCAAT-box in plant promoters	100,0	98,3
9	59	271PSCAAT.01	CCAAT-box in plant promoters	100,0	97,9
9	62	269PSCARICH.01	CA-rich element	100,0	79,7
9	63	284PSCCA1.01	Circadian clock associated 1	100,0	91,1
9	68	270PSDOF1.01	Dof1 / MNB1a - single zinc finger transcription factor	100,0	99,6
9	70	277PSDOF3.01	Dof3 - single zinc finger transcription factor	100,0	99,2
9	80	282PSGAP.01	Cis-element in the GAPDH promoters conferring light inducibility	100,0	93,3
9	80	284PSGAP.01	Cis-element in the GAPDH promoters conferring light inducibility	100,0	91,5
9	80	305PSGAP.01	Cis-element in the GAPDH promoters conferring light inducibility	100,0	88,4
9	81	280PSGATA.01	Class I GATA factors	100,0	97,0

			GCN4, conserved in cereal seed storage protein gene promoters, similar to yeast GCN4 and vertebrate AP-1	100,0	89,5
9	84	308 P$GCN4.01	GCN4, conserved in cereal seed storage protein gene promoters, similar to yeast GCN4 and vertebrate AP-1	100,0	83,2
0	84	310 PSGCN4.01	GCN4, conserved in cereal seed storage protein gene promoters, similar to yeast GCN4 and vertebrate AP-1	100,0	83,2
0	85	270 PSGT1.01	GT1-Box binding factors with a trihelix DNA-binding domain	99,8	85,2
0	87	278 PSHAHB4.01	Sunflower homeodomain leucine-zipper protein Hahb-4	100,0	91,3
0	96	281 PSICE.01	ICE (inducer of CBF expression 1), AtMYC2 (rd22BP1)	100,0	96,4
0	97	282 PSID1.01	Maize INDETERMINATE1 zinc finger protein	100,0	96,2
0	100	289 PSLFY.01	Plant specific floral meristem identity gene LEAFY (LFY)	100,0	95,8
0	105	300 PSMYBPH3.02	Myb-like protein of Petunia hybrida	91,7	94,1
0	106	292 PSMYBST1.01	MybSt1 (Myb Solanum tuberosum 1) with a single	100,0	96,9

			myb repeat			
9	108	275PSNCS1.01	Nodulin consensus sequence 1	100.0	100.0	90.3
9	108	285PSNCS1.01	Nodulin consensus sequence 1	100.0	100.0	99.0
9	108	287PSNCS1.01	Nodulin consensus sequence 1	100.0	100.0	85.3
9	109	290PSNCS2.01	Nodulin consensus sequence 2	76.0	76.0	86.1
9	114	293PSO2_CON4.01	Recognition site for BZIP transcription factors that belong to the group of Opaque-2 like proteins	76.6	76.6	82.0
9	115	280PSOCSL.01	OCS-like elements	80.7	80.7	70.7
9	115	286PSOCSL.01	OCS-like elements	80.7	80.7	74.6
9	124	261PSPBOX.01	Prolamin box, conserved in cereal seed storage protein gene promoters	76.1	76.1	76.9
9	124	314PSPBOX.01	Prolamin box, conserved in cereal seed storage protein gene promoters	100.0	100.0	79.3
9	125	268PSPCF2.01	TCP class I transcription factor	100.0	100.0	99.1

9	135	272 P$S1F.01	S1F, site 1 binding factor of spinach rps1 promoter	100,0	79,8
9	136	201 PS1F.01	S1F, site 1 binding factor of *spinach rps1 promoter*	100,0	81,1
9	130	307 PSBF1.01	SBF-1	100,0	88,7
9	139	312 PSBF1.01	SBF-1	100,0	91,1
9	141	297 PSQUA.01	MADS-box protein SQUAMOSA	100,0	90,2
9	141	311 PSQUA.01	MADS-box protein SQUAMOSA	100,0	92,5
9	142	276 PSSTK.01	Storekeeper (STK), plant specific DNA binding protein important for tuber-specific and sucrose-inducible gene expression	100,0	91,0
9	142	286 PSSTK.01	Storekeeper (STK), plant specific DNA binding protein important for tuber-specific and sucrose-inducible gene expression	100,0	88,0
9	142	302 PSSTK.01	Storekeeper (STK), plant specific DNA binding protein important for tuber-specific	100,0	87,0

and sucrose-inducible gene expression							
			0	142	30|PSSTK.01	Storekeeper (STK) plant specific DNA binding protein important for tuber-specific and sucrose-inducible gene expression	100.0	86.4
			0	144	267|PSTAMY600.01	MYB protein from wheat	100.0	86.0
			0	140	269|PSTATA.01	Plant TATA box	100.0	90.7
			0	140	274|PSTATA.01	Plant TATA box	100.0	91.6
			0	140	309|PSTATA.01	Plant TATA box	100.0	88.0
			0	148	280|PSTEF1.01	TEF cis acting elements in both RNA polymerase II-dependent promoters and rDNA spacer sequences	100.0	78.2
			0	149	315|PSTEIL.01	TEIL (tobacco EIN3-like)	98.4	92.6
			0	152	209|PSUSE.01	Upstream sequence elements in the promoters of U-snRNA genes of higher plants	100.0	70.3
			0	154	280|PSWUS.01	Homeodomain protein WUSCHEL	100.0	100.0
			0	154	279|PSWUS.01	Homeodomain protein	100.0	100.0

WUSCHEL									
	Homeodomain protein WUSCHEL	284PSWUS.01	154	9	100,0	100,0	66,3		
	ABA (abscisic acid) inducible transcriptional activator	620PSABF1.01	28	10	100,0	100,0	80,6		
	ABA (abscisic acid) inducible transcriptional activator	601PSABF1.03	28	10	100,0	100,0	88,0		
	ABA response elements	606PSABRE.01	30	10	100,0	100,0	82,0		
	AGL2, Arabidopsis MADS-domain protein AGAMOUS-like 2	578PSAGL2.01	35	10	70,3	99,2	83,4		
	AGL3, MADS Box protein	560PSAGL3.02	37	10	99,2	100,0	84,8		
	Arabidopsis thaliana homeo box protein 1	570PSATHB1.01	45	10	100,0	78,9	98,9		
	Arabidopsis thaliana homeo box protein 1	571PSATHB1.01	45	10	78,9	100,0	90,0		
	Arabidopsis thaliana homeo box protein 1	566PSATHB1.01	45	10	100,0	78,9	99,0		
	Arabidopsis thaliana homeo box protein 1	616PSATHB1.01	45	10	78,9	82,9	90,1		
	HDZip class I protein ATHB5	587PSATHB5.01	40	10	82,9	93,6	94,0		
	HDZip class I protein ATHB5	617PSATHB5.01	40	10	93,6	100,0	93,6		

			L1-specific homeodomain protein ATML1 (A. thaliana meristem layer 1)	75,0		83,8
10	48	573 P$ATML1.01	L1-specific homeodomain protein ATML1 (A. thaliana meristem layer 1)	75,0		82,3
10	48	631 PSATML1.01		75,0		82,3
10	61	507 PSCARE.01	CAACTC regulatory elements, CA-inducible	100,0		83,0
10	62	509 PSCARICH.01	CA-rich element	100,0		81,6
10	62	560 PSCARICH.01	CA-rich element	100,0		81,2
10	62	561 PSCARICH.01	CA-rich element	100,0		78,3
10	63	560 PSCCA1.01	Circadian clock associated 1	100,0		87,3
10	66	586 PSCPRF.01	Common plant regulatory factor (CPRF) from parsley	100,0		99,3
10	66	597 PSCPRF.01	Common plant regulatory factor (CPRF) from parsley	100,0		93,7
10	69	579 PSDOF1.01	Dof1 / MNB1a - single zinc finger transcription factor	100,0		93,4
10	77	577 PSCAAA.01	CAAA motif involved in pollen specific transcriptional activation	100,0		60,3
10	79	559 PSCAMYB.01	CA-regulated myb gene from barley	100,0		91,9

10	80	623P$GAP.01	Cis-element in the GAPDH promoters conferring light inducibility	100,0	95,0
10	81	589PSGATA.01	Class I GATA factors	100,0	96,7
10	86	604PSGT1.01	GT1-Box binding factors with a trihelix DNA-binding domain	96,6	86,3
10	86	606PSGT1.01	GT1-Box binding factors with a trihelix DNA-binding domain	96,6	87,5
10	86	606PSGT3A.01	Trihelix DNA-binding factor GT-3a	100,0	83,9
10	86	607PSGT3A.01	Trihelix DNA-binding factor GT-3a	100,0	90,2
10	87	600PSHAHB4.01	Sunflower homeodomain leucine-zipper protein Hahb-4	100,0	94,1
10	89	594PSHBP1B.01	Wheat bZIP transcription factor HBP1b (histone gene binding protein 1b)	100,0	89,5
10	97	572PSID1.01	Maize INDETERMINATE1 zinc finger protein	100,0	96,2
10	97	603PSID1.01	Maize INDETERMINATE1 zinc finger protein	100,0	96,4

			Legumin box, highly conserved sequence element about 100 bp upstream of the TSS in legumin genes	75,0		63,2
10	99	629P$LEGB.01				
10	105	627PSMYBPH3.02	Myb-like protein of Petunia hybrida	100,0		77,3
10	106	599PSMYBST1.01	MybSt1 (Myb Solanum tuberosum 1) with a single myb repeat	100,0		92,7
10	107	602PSMYORS.01	Myc recognition sequences	100,0		94,3
10	107	621ESMYORS.01	Myc recognition sequences	86,3		93,3
10	108	616PSNCS1.01	Nodulin consensus sequence 1	87,8		91,0
10	114	630PSO2_GCN4.01	Recognition site for BZIP transcription factors that belong to the group of Opaque-2 like proteins	100,0		88,9
10	115	670PSOCSL.01	OCS-like elements	80,7		69,6
10	119	583PSOSMYBS.01	Rice MYB proteins with single DNA binding domains, binding to the amylase element (TATCCA)	75,0		83,7

			Rice MYB proteins with single DNA binding domains, binding to the amylase element (TATCCA)	75,0	82,8
10	119	612P$OSMYBS.01	Rice MYB proteins with single DNA binding domains, binding to the amylase element (TATCCA)	75,0	82,0
10	119	628PSOSMYBS.01	Prolamin box, conserved in cereal seed storage protein gene promoters	77,6	77,2
10	124	574PSPBOX.01	Prolamin box, conserved in cereal seed storage protein gene promoters	77,6	76,9
10	124	575PSPBOX.01	Prolamin box, conserved in cereal seed storage protein gene promoters	76,1	76,0
10	124	608PSPBOX.01	Prolamin box, conserved in cereal seed storage protein gene promoters	100,0	77,8
10	124	614PSPBOX.01	Prolamin box, conserved in cereal seed storage protein gene promoters	77,6	76,8
10	124	624PSPBOX.01			

			Rice transcription activator-1 (RITA), basic leucin zipper protein, highly expressed during seed development	100,0		96,6
10	131	600 P$RITA1.01				
10	133	556 PSRPBX.01	Ribosomal protein box, appears unique to plant RP genes and genes associated with gene expression	100,0		85,2
10	134	582 PSRY.01	RY and Sph motifs conserved in seed-specific promoters	100,0		91,7
10	134	581 PSRY.01	RY and Sph motifs conserved in seed-specific promoters	100,0		88,0
10	134	608 PSRY.01	RY and Sph motifs conserved in seed-specific promoters	100,0		90,0
10	134	609 PSRY.01	RY and Sph motifs conserved in seed-specific promoters	100,0		99,0
10	134	610 PSRY.01	RY and Sph motifs conserved in seed-specific promoters	100,0		96,7
10	134	616 PSRY.01	RY and Sph motifs conserved in seed-specific promoters	100,0		92,8
10	134	625 PSRY.01	RY and Sph motifs conserved in seed-specific promoters	100,0		88,4

10	134	626 P$RY.01	RY and Sph motifs conserved in seed-specific promoters	100,0	92,8
10	136	564PSS1F.01	S1F, site 1 binding factor of spinach rps1 promoter	75,0	79,4
10	136	616PSS1F.01	S1F, site 1 binding factor of spinach rps1 promoter	100,0	80,0
10	136	622PSS1F.01	S1F, site 1 binding factor of spinach rps1 promoter	100,0	79,7
10	139	557PSSEF4.01	Soybean embryo factor 4	100,0	98,2
10	140	559PSSP6BF.01	DNA-binding protein of sweet potato that binds to the SP6a (ACTGTGTA) and SP8b (TACTATT) sequences of sporamin and beta-amylase genes	100,0	90,1
10	142	569PSSTK.01	Storekeeper (STK) plant specific DNA binding protein important for tuber-specific and sucrose-inducible gene expression	100,0	88,2
10	144	594PSTAMY680.01	MYB protein from wheat	100,0	94,4
10	144	585PSTAMY680.01	MYB protein from wheat	100,0	92,3

10	145	599	P$TANAC69.01	Wheat NACdomain DNA binding factor	89,5	71,1
10	147	561	PSTATA.02	Plant TATA box	100,0	93,1
10	147	562	PSTATA.02	Plant TATA box	100,0	90,7
10	149	592	PSTEIL.01	TEIL (tobacco EIN3-like)	89,5	97,4
10	149	566	PSTEIL.01	TEIL (tobacco EIN3-like)	96,4	93,2
10	149	611	PSTEIL.01	TEIL (tobacco EIN3-like)	89,5	97,4
10	150	593	PSTGA1.01	Arabidopsis leucine zipper protein TGA1	100,0	96,8
10	152	559	PSUSE.01	Upstream sequence elements in the promoters of U-snRNA genes of higher plants	75,0	70,4
10	152	613	PSUSE.01	Upstream sequence elements in the promoters of U-snRNA genes of higher plants	100,0	76,4
10	154	568	PSWUS.01	Homeodomain protein WUSCHEL	100,0	96,3
11	29	550	PSABI4.01	ABA insensitive protein 4 (ABI4) (AtABI4 - Arabidopsis, ZmABI4 - maize, OsABI4 - rice)	100,0	100,0
11	30	545	PSABRE.01	ABA response elements	100,0	90,0
11	30	540	PSABRE.01	ABA response elements	100,0	87,9

			AGL15, Arabidopsis MADS-domain protein AGAMOUS-like 15	85,0	80,7
11	34	522 P$AGL15.01			
11	35	512PSAGL2.01	AGL2, Arabidopsis MADS-domain protein AGAMOUS-like 2	70,5	86,2
11	36	518PSAGL3.01	AGL3, MADS Box protein	97,3	89,2
11	36	532PSAGL3.01	AGL3, MADS Box protein	85,3	85,1
11	43	529PSARR10.01	Type-B response regulator (ARR10), member of the GARP-family of plant myb-related DNA binding motifs	100,0	97,1
11	44	510PSATCTA.01	Motif involved in carotenoid and tocopherol biosynthesis and in the expression of photosynthesis-related genes	100,0	97,5
11	44	528PSATCTA.01	Motif involved in carotenoid and tocopherol biosynthesis and in the expression of photosynthesis-related genes	100,0	97,0
11	44	536PSATCTA.01	Motif involved in carotenoid and tocopherol biosynthesis and in the expression of	100,0	89,3

			photosynthesis-related genes		
11	46	519PSATHB5.01	HDZip class I protein ATHB5	82.9	94.0
11	46	520PSATHB5.01	HDZip class I protein ATHB5	100.0	100.0
11	48	524PSATML1.01	L1-specific homeodomain protein ATML1 (A. thaliana meristem layer 1)	100.0	87.3
11	49	527PSATML1.01	L1-specific homeodomain protein ATML1 (A. thaliana meristem layer 1)	75.0	82.4
11	52	506PSATPURA.01	Arabidopsis Tele-box interacting protein related to the conserved animal protein Pur-alpha	100.0	88.1
11	57	551PSBZIP01.01	bZIP transcription factor from Antirrhinum majus	75.0	78.3
11	57	555PSBZIP01.01	bZIP transcription factor from Antirrhinum majus	75.0	77.1
11	63	522PSCCA1.01	Circadian clock associated 1	100.0	85.9
11	66	529PSDOF1.01	Dof1 / MNB1a - single zinc finger transcription factor	100.0	99.6
11	72	504PSE2F.01	E2F class I sites	100.0	82.0

			GAAA motif involved in pollen specific transcriptional activation	100,0	85,2
11	77	503 P$GAAA.01	bZIP protein G-Box binding factor 1	100,0	90,7
11	82	530 PSGBF1.01	bZIP protein G-Box binding factor 1	100,0	97,7
11	82	545 PSGBF1.01	bZIP protein G-Box binding factor 1	100,0	97,6
11	82	544 PSGBF1.01	GT1-Box binding factors with a trihelix DNA-binding domain	100,0	92,0
11	86	521 PSGT1.01	Trihelix DNA-binding factor GT-3a	100,0	84,7
11	86	546 PSGT3A.01	Sunflower homeodomain leucine-zipper protein Hahb-4	100,0	90,1
11	87	537 PSHAHB4.01	Wheat bZIP transcription factor HBP1B (histone gene binding protein 1b)	78,9	84,5
11	89	516 PSHBP1B.01	High mobility group I/Y-like proteins	100,0	89,3
11	90	513 PSHMG_IY.01	ICE (inducer of CBF expression 1), AtMYC2 (rd22BP1)	100,0	97,0
11	96	539 PSICE.01			

11	96	547 P$ICE.01	ICE (inducer of CBF expression 1), AtMYC2 (rd22BP1)	95,4	96,6
11	99	540 PSLEGB.01	Legumin box, highly conserved sequence element about 100 bp upstream of the TSS in legumin genes	100,0	100,0
11	104	510 PSMYBPH3.01	Myb-like protein of Petunia hybrida	100,0	88,6
11	106	508 PSMYBST1.01	MybSt1 (Myb Solanum tuberosum (1) with a single myb repeat	100,0	93,2
11	109	533 PSNCS2.01	Nodulin consensus sequence 2	100,0	88,5
11	113	552 PSO2.02	Opaque-2 regulatory protein	100,0	87,0
11	114	541 PSOCSL.01	OCS-like elements	100,0	70,4
11	119	539 PSOSMYBS.01	Rice MYB proteins with single DNA binding domains, binding to the amylase element (TATCCA)	100,0	91,3
11	124	526 PSP1BOX.01	Prolamin box, conserved in cereal seed storage protein gene promoters	100,0	86,9

			5'-part of bipartite RAV1 binding site, interacting with AP2 domain	100,0	98,7
11	130	531 P$RAV1-5.01			
11	134	529 PSRY.01	RY and Sph motifs conserved in seed-specific promoters	100,0	94,5
11	136	511 PSSBF1.01	SBF-1	100,0	88,6
11	136	517 PSSBF1.01	SBF-1	100,0	88,6
11	137	542 PSSEBF.01	Silencing element binding factor - transcriptional repressor	100,0	97,6
11	139	514 PSSEF4.01	Soybean embryo factor 4	100,0	98,7
11	140	509 PSSP6BF.01	DNA-binding protein of sweet potato that binds to the SP8a (ACTCTCTA) and SP8b (TACTATT) sequences of sporamin and beta-amylase genes	100,0	93,9
11	140	507 PSSPSBF.01	DNA-binding protein of sweet potato that binds to the SP8a (ACTCTCTA) and SP8b (TACTATT) sequences of sporamin and beta-amylase genes	81,4	90,1

DNA-binding protein of sweet potato that binds to the SP8a (ACTGTGTA) and SP8b (TACTATT) sequences of sporamin and beta-amylase genes					100,0	93,5
	509	P$SP8BF.01				
	140					
			DNA-binding protein of sweet potato that binds to the SP8a (ACTGTGTA) and SP8b (TACTATT) sequences of sporamin and beta-amylase genes		100,0	87,1
			554PSSP8BF.01			
	140					
11						
			MYB protein from wheat		100,0	85,8
			554PSTAMYB80.01			
	144					
11						
			Plant TATA box		100,0	94,0
			549PSTATA.02			
	147					
11						
			ABA (abscisic acid) inducible transcriptional activator		100,0	83,5
			382PSABF1.01			
	20					
12						
			ABA insensitive protein 4 (ABI4) (AtABI4 - Arabidopsis, ZmABI4 - maize, OsABI4 - rice)		100,0	100,0
			498PSABI4.01			
	29					
12						
			ABA response elements		100,0	90,0
			469PSABRE.01			
	30					
12						
			ABA response elements		100,0	87,9
			494PSABRE.01			
	30					
12						

12	33	350 P$AGL1.01	AGL1, Arabidopsis MADS-domain protein AGAMOUS-like 1	91,5		88,6
12	33	351 PSAGL1.01	AGL1, Arabidopsis MADS-domain protein AGAMOUS-like 1	100,0		91,8
12	34	356 PSAGL15.01	AGL15, Arabidopsis MADS-domain protein AGAMOUS-like 15	100,0		81,7
12	34	470 PSAGL15.01	AGL15, Arabidopsis MADS-domain protein AGAMOUS-like 15	85,0		80,7
12	35	469 PSAGL2.01	AGL2, Arabidopsis MADS-domain protein AGAMOUS-like 2	79,5		86,2
12	36	372 PSAGL3.01	AGL3, MADS Box protein	97,3		85,6
12	36	408 PSAGL3.01	AGL3, MADS Box protein	97,3		83,2
12	36	480 PSAGL3.01	AGL3, MADS Box protein	85,3		85,1
12	42	342 PSARE.01	Auxin Response Element	100,0		93,2
12	42	360 PSARE.01	Auxin Response Element	100,0		93,2
12	42	360 PSARE.01	Auxin Response Element	100,0		97,0

			Type-B response regulator (ARR10), member of the GARP-family of plant myb-related DNA binding motifs	100,0		97,1
12	43	477	P$ARR10.01			
12	44	377	PSATCTA.01	Motif involved in carotenoid and tocopherol biosynthesis and in the expression of photosynthesis-related genes	100,0	91,0
12	44	383	PSATCTA.01	Motif involved in carotenoid and tocopherol biosynthesis and in the expression of photosynthesis-related genes	100,0	89,2
12	44	388	PSATCTA.01	Motif involved in carotenoid and tocopherol biosynthesis and in the expression of photosynthesis-related genes	100,0	90,7
12	44	423	PSATCTA.01	Motif involved in carotenoid and tocopherol biosynthesis and in the expression of photosynthesis-related genes	100,0	87,2
12	44	464	PSATCTA.01	Motif involved in carotenoid and tocopherol biosynthesis and in the expression of	100,0	97,5

							97.0
12		44	479PSATCTA.01	photosynthesis-related genes	Motif involved in carotenoid and tocopherol biosynthesis and in the expression of	100.0	
12		44	484PSATCTA.01	photosynthesis-related genes	Motif involved in carotenoid and tocopherol biosynthesis and in the expression of	100.0	89.3
12		45	337PSATHB1.01	box protein 1	Arabidopsis thaliana homeo	100.0	98.6
12		45	338PSATHB1.01	box protein 1	Arabidopsis thaliana homeo	78.9	90.1
12		46	339PSATHB1.01	box protein 1	Arabidopsis thaliana homeo	100.0	98.6
12		46	340PSATHB5.01	HDZip class I protein ATHB5		82.9	94.0
12		46	302PSATHB5.01	HDZip class I protein ATHB5		93.6	93.6
12		46	303PSATHB5.01	HDZip class I protein ATHB5		82.9	94.0
12		46	309PSATHB5.01	HDZip class I protein ATHB5		93.6	97.7
12		46	306PSATHB5.01	HDZip class I protein ATHB5		82.9	94.0
12		46	467PSATHB5.01	HDZip class I protein ATHB5		82.9	94.0

12	46	468	P$ATHB5.01	HDZip class I protein ATHB5	100,0	100,0
12	47	329	PSATHB9.01	HD-ZIP class III protein ATHB9	100,0	89,1
12	46	368	PSATML1.01	L1-specific homeodomain protein ATML1 (A. thaliana meristem layer 1)	76,0	83,0
12	48	389	PSATML1.01	L1-specific homeodomain protein ATML1 (A. thaliana meristem layer 1)	100,0	88,8
12	48	430	PSATML1.01	L1-specific homeodomain protein ATML1 (A. thaliana meristem layer 1)	100,0	80,3
12	48	459	PSATML1.01	L1-specific homeodomain protein ATML1 (A. thaliana meristem layer 1)	76,0	83,4
12	48	472	PSATML1.01	L1-specific homeodomain protein ATML1 (A. thaliana meristem layer 1)	100,0	87,3
12	48	475	PSATML1.01	L1-specific homeodomain protein ATML1 (A. thaliana meristem layer 1)	76,0	82,5
12	50	360	PSATMYB77.01	R2R3-type myb-like transcription factor (H-type	100,0	92,9

							95.5
						100.0	
12	50	363PSATMYB77.01	RZR3-type myb-like transcription factor (I-type binding site)	100.0			
							89.4
						100.0	
12	50	361PSATMYB77.01	RZR3-type myb-like transcription factor (I-type binding site)	100.0			
							87.4
						86.7	
12	50	339PSATMYB77.01	RZR3-type myb-like transcription factor (I-type binding site)	86.7			
							88.4
						100.0	
12	52	332PSATPURA.01	Arabidopsis Telo-box interacting protein related to the conserved animal protein Pur-alpha	100.0			
							86.9
						100.0	
12	52	436PSATPURA.01	Arabidopsis Telo-box interacting protein related to the conserved animal protein Pur-alpha	100.0			
							88.1
						100.0	
12	52	453PSATPURA.01	Arabidopsis Telo-box interacting protein related to the conserved animal protein Pur-alpha	100.0			

			Arabidopsis thaliana signal-responsive gene1, Ca2+/ calmodulin binding protein homolog to NtER1 (tobacco early ethylene-responsive gene)	100,0		88,6
12	53	401P$ATSR1.01	Arabidopsis thaliana signal-responsive gene1, Ca2+/ calmodulin binding protein homolog to NtER1 (tobacco early ethylene-responsive gene)	100,0		89,0
12	59	402PSATSR1.01		100,0		89,0
12	57	499PSBZIP911.01	bZIP transcription factor from Antirrhinum majus	75,0		78,3
12	57	501PSBZIP911.01	bZIP transcription factor from Antirrhinum majus	75,0		77,1
12	59	439PSCAAT.01	CCAAT-box in plant promoters	100,0		98,3
12	59	440PSCAAT.01	CCAAT-box in plant promoters	100,0		97,9
12	63	416PSCCA1.01	Circadian clock associated 1	100,0		88,8
12	63	432PSCCA1.01	Circadian clock associated 1	100,0		88,0
12	63	471PSCCA1.01	Circadian clock associated 1	100,0		85,9
12	64	400PSCE3.01	Coupling element 3 (CE3), non-ACGT ABRE	100,0		80,6

12	68	373 P$DOF1.01	Dof1 / MNB1a - single zinc finger transcription factor	100,0	98,7
12	68	476PSDOF1.01	Dof1 / MNB1a - single zinc finger transcription factor	100,0	99,6
12	70	459PSDOF3.01	Dof3 - single zinc finger transcription factor	100,0	99,4
12	72	451PSE2F.01	E2F class I sites	100,0	82,0
12	74	352PSERE.01	Elicitor response element	100,0	89,9
12	76	355PSERSE.I.01	ERSE I (ER stress-response element I)-like motif	100,0	83,5
12	77	336PSGAAA.01	GAAA motif involved in pollen specific transcriptional activation	100,0	83,3
12	79	317PSGAMYB.01	GA-regulated myb gene from barley	100,0	93,7
12	79	410PSGAMYB.01	GA-regulated myb gene from barley	100,0	92,2
12	81	349PSGATA.01	Class I GATA factors	100,0	97,1
12	81	384PSGATA.01	Class I GATA factors	100,0	93,9
12	81	391PSGATA.01	Class I GATA factors	100,0	96,4
12	81	409PSGATA.01	Class I GATA factors	100,0	94,0
12	81	428PSGATA.01	Class I GATA factors	100,0	94,4

			bZIP protein G-Box binding factor 1	100,0	96,7
12	82	478 P$GBF1.01	bZIP protein G-Box binding factor 1	100,0	97,7
		491 PSGBF1.01			
12	82				
12	82	492 PSGBF1.01	bZIP protein G-Box binding factor 1	100,0	97,6
12	64	375 PSGCN4.01	GCN4, conserved in cereal seed storage protein gene promoters, similar to yeast GCN4 and vertebrate AP-1	76,1	89,3
12	86	327 PSGT1.01	GT1-Box binding factors with a trihelix DNA-binding domain	84,3	80,8
12	89	345 PSGT1.01	GT1-Box binding factors with a trihelix DNA-binding domain	100,0	85,6
12	86	307 PSGT1.01	GT1-Box binding factors with a trihelix DNA-binding domain	96,8	88,9
12	85	427 PSGT1.01	GT1-Box binding factors with a trihelix DNA-binding domain	96,8	87,0
12	86	445 PSGT1.01	GT1-Box binding factors with a trihelix DNA-binding domain	84,3	89,2
12	86	468 PSGT1.01	GT1-Box binding factors with a trihelix DNA-binding domain	100,0	92,0
12	86	320 PSGT3A.01	Trihelix DNA-binding factor	100,0	87,6

GT-3a						
12	86	387PSGT3A.01	Trihelix DNA-binding factor GT-3a	100.0	100.0	84.7
12	86	498PSGT3A.01	Trihelix DNA-binding factor GT-3a	100.0	100.0	84.7
12	87	419PSHAHB4.01	Sunflower homeodomain leucine-zipper protein Hahb-4	100.0	100.0	91.6
12	87	420PSHAHB4.01	Sunflower homeodomain leucine-zipper protein Hahb-4	100.0	100.0	91.0
12	87	462PSHAHB4.01	Sunflower homeodomain leucine-zipper protein Hahb-4	100.0	100.0	90.1
12	89	396PSHBP1B.01	Wheat bZIP transcription factor HBP1B (histone gene binding protein 1b)	76.9	76.9	83.3
12	89	438PSHBP1B.01	Wheat bZIP transcription factor HBP1B (histone gene binding protein 1b)	76.9	76.9	85.1
12	89	462PSHBP1B.01	Wheat bZIP transcription factor HBP1B (histone gene binding protein 1b)	76.9	76.9	84.5
12	90	339PSHMG_IY.01	High mobility group I/Y-like proteins	100.0	100.0	89.8

12	90	460 P$HMG_IY.01	High mobility group I/Y-like proteins	100,0	91,5
12	92	316PSHSE.01	Heat shock element	78.2	87.0
12	92	424PSHSE.01	Heat shock element	100.0	85.0
12	96	468PSICE.01	ICE (inducer of CBF expression 1), AtMYC2 (rd22BP1)	100.0	97.0
12	96	469PSICE.01	ICE (inducer of CBF expression 1), AtMYC2 (rd22BP1)	96.4	96.6
12	97	324PSID1.01	Maize INDETERMINATE1 zinc finger protein	100.0	97.0
12	97	341PSID1.01	Maize INDETERMINATE1 zinc finger protein	100.0	92.3
12	97	339PSID1.01	Maize INDETERMINATE1 zinc finger protein	100.0	92.2
12	97	304PSID1.01	Maize INDETERMINATE1 zinc finger protein	100.0	92.9
12	97	412PSID1.01	Maize INDETERMINATE1 zinc finger protein	100.0	94.1
12	99	468PSLEGB.01	Legumin box, highly conserved sequence element about 100 bp upstream of the	100.0	100.0

TSS in legumin genes							
Binding sites for AP1, AP3-PI and AG dimers	419PSMADS.01	101	12	84,6	84,6	78,0	
Binding sites for AP1, AP3-PI and AG dimers	417PSMADS.01	101	12	84,6	84,6	88,7	
Myb-like protein of Petunia hybrida	376PSMYBPH3.01	104	12	100,0	100,0	84,7	
Myb-like protein of Petunia hybrida	457PSMYBPH3.01	104	12	100,0	100,0	88,8	
Myb-like protein of Petunia hybrida	322PSMYBPH3.02	105	12	100,0	100,0	81,5	
Myb-like protein of Petunia hybrida	329PSMYBPH3.02	105	12	81,7	81,7	79,8	
Myb-like protein of Petunia hybrida	340PSMYBPH3.02	106	12	100,0	100,0	82,0	
Myb-like protein of Petunia hybrida	354PSMYBPH3.02	105	12	100,0	100,0	77,6	
Myb-like protein of Petunia hybrida	359PSMYBPH3.02	106	12	77,8	77,8	82,7	
Myb-like protein of Petunia hybrida	409PSMYBPH3.02	105	12	100,0	100,0	79,2	

			MybSt1 (Myb Solanum tuberosum 1) with a single myb repeat	100,0	93,2
12	106	455 P$MYBST1.01	Nodulin consensus sequence 2	100,0	80,6
12	109	374 PSNCS2.01			
12	109	415 PSNCS2.01	Nodulin consensus sequence 2	76,6	81,4
12	109	481 PSNCS2.01	Nodulin consensus sequence 2	100,0	88,5
12	110	379 PSNCS3.01	Nodulin consensus sequence 3	100,0	91,8
12	111	305 PSNTLIM1.01	LIM domain protein binding to a PAL-box like sequence	100,0	90,9
12	112	510 PSO2.01	Opaque-2 regulatory protein	100,0	90,1
12	112	369 PSO2.01	Opaque-2 regulatory protein	100,0	92,3
12	112	404 PSO2.01	Opaque-2 regulatory protein	100,0	90,1
12	113	500 PSO2.02	Opaque-2 regulatory protein	100,0	87,0
12	114	370 PSO2_GCN4.01	Recognition site for BZIP transcription factors that belong to the group of Opaque-2 like proteins	100,0	81,7

			Recognition site for BZIP transcription factors that belong to the group of Opaque-2 like proteins	82,9		84,6
12	114	443P$O2_GCN4.01				
12	115	359PSOCSL01	OCS-like elements	70,9		73,8
12	115	361PSOCSL01	OCS-like elements	70,9		69,3
12	115	421PSOCSL01	OCS-like elements	60,7		69,7
12	115	426PSOCSL01	OCS-like elements	80,7		69,6
12	115	447PSOCSL01	OCS-like elements	70,9		72,6
12	115	469PSOCSL01	OCS-like elements	100,0		70,4
12	119	367PSOSMYBS01	Rice MYB proteins with single DNA binding domains, binding to the amylase element (TATCCA)	100,0		94,2
12	119	483PSOSMYBS01	Rice MYB proteins with single DNA binding domains, binding to the amylase element (TATCCA)	100,0		91,3
12	120	364PSPALBOXA01	Putative cis-acting element on various PAL and 4CL gene promoters	100,0		90,3
12	123	414PSPBF01	PBF (MPBF)	100,0		97,0

			Prolamin box, conserved in cereal seed storage protein gene promoters	76,1		78,5
12	124	343 P$PBOX.01				
12	124	344PSPBOX.01	Prolamin box, conserved in cereal seed storage protein gene promoters	100,0	12	78,4
12	124	347PSPBOX.01	Prolamin box, conserved in cereal seed storage protein gene promoters	100,0	12	86,7
12	124	449PSPBOX.01	Prolamin box, conserved in cereal seed storage protein gene promoters	100,0	12	82,3
12	124	474PSPBOX.01	Prolamin box, conserved in cereal seed storage protein gene promoters	100,0	12	86,9
12	126	329PSP_ACT.01	Maize activator P of flavonoid biosynthetic genes	100,0	12	93,2
12	130	479PSRAV1-5.01	5'-part of bipartite RAV1 binding site, interacting with AP2 domain	100,0	12	98,7
12	139	571PSRPBX.01	Ribosomal protein box, appears unique to plant RP genes and genes associated	76,5	12	87,1

			with gene expression			
			Ribosomal protein box, appears unique to plant RP genes and genes associated with gene expression			
12	133	440PSRPBX.01		100,0	100,0	84,6
			RY and Sph motifs conserved in seed-specific promoters			
12	134	357PSRY.01		100,0	100,0	90,2
			RY and Sph motifs conserved in seed-specific promoters			
12	134	487PSRY.01		100,0	100,0	94,6
			S.I.F. site 1 binding factor of spinach rps1 promoter			
12	135	365PSSIF.01		100,0	100,0	79,7
12	136	430PSSBF1.01	SBF-1	100,0	100,0	90,6
12	136	442PSSBF1.01	SBF-1	100,0	100,0	88,6
12	136	459PSSBF1.01	SBF-1	100,0	100,0	88,6
12	136	463PSSBF1.01	SBF-1	100,0	100,0	88,0
			Silencing element binding factor - transcriptional repressor			
12	137	490PSSEBF.01		100,0	100,0	97,6
			Soybean embryo factor 4			
12	139	462PSSEF4.01		100,0	100,0	98,7

									91,5
									100,0
DNA-binding protein of sweet potato that binds to the SP8a (ACTGTGTA) and SP8b (TACTATT) sequences of sporamin and beta-amylase genes									
	12	140	411	P$SP8BF.01					
DNA-binding protein of sweet potato that binds to the SP8a (ACTGTGTA) and SP8b (TACTATT) sequences of sporamin and beta-amylase genes	12	140	429	PSP8BF.01					88,6
DNA-binding protein of sweet potato that binds to the SP8a (ACTGTGTA) and SP8b (TACTATT) sequences of sporamin and beta-amylase genes	12	140	452	PSP8BF.01					93,9
DNA-binding protein of sweet potato that binds to the SP8a (ACTGTGTA) and SP8b (TACTATT) sequences of sporamin and beta-amylase genes	12	140	454	PSP8BF.01				81,4	90,1

genes							
DNA-binding protein of sweet potato that binds to the SPFa (ACTCTGTA) and SPFb (TACTATT) sequences of sporamin and beta-amylase genes	45PSSP8BF.01	140	12	140	100.0	93.3	
DNA-binding protein of sweet potato that binds to the SPFa (ACTCTGTA) and SPFb (TACTATT) sequences of sporamin and beta-amylase genes	50PSSP8BF.01	140	12	100.0	87.1		
MADS-box protein SOLAMOSA	30PSSOLA.01	141	12	100.0	92.8		
Storekeeper (STK), plant specific DNA binding protein important for tuber-specific and sucrose-inducible gene expression	40PSSTK.01	142	12	78.6	85.1		

				Storekeeper (STK), plant specific DNA binding protein important for tuber-specific and sucrose-inducible gene expression	100,0		89,7
12	142	425	P$STK.01				
12	142	461	PSSTK.01	Storekeeper (STK), plant specific DNA binding protein important for tuber-specific and sucrose-inducible gene expression	100,0	85,9	85,9
12	144	482	PSTAMYB00.01	MYB protein from wheat	100,0	85,6	85,6
12	145	340	PSTANAC00.01	Wheat NAC/domain DNA binding factor	100,0	70,1	70,1
12	140	310	PSTATA.01	Plant TATA box	100,0	91,6	91,6
12	140	409	PSTATA.01	Plant TATA box	100,0	94,5	94,5
12	140	408	PSTATA.01	Plant TATA box	100,0	88,6	88,6
12	140	431	PSTATA.01	Plant TATA box	82,6	88,3	88,3
12	147	331	PSTATA.02	Plant TATA box	100,0	96,0	96,0
12	147	335	PSTATA.02	Plant TATA box	100,0	90,2	90,2
12	147	407	PSTATA.02	Plant TATA box	100,0	93,9	93,9
12	147	418	PSTATA.02	Plant TATA box	100,0	93,6	93,6
12	147	441	PSTATA.02	Plant TATA box	100,0	95,3	95,3
12	147	497	PSTATA.02	Plant TATA box	100,0	94,0	94,0

			TEF cis acting elements in both RNA polymerase II-dependent promoters and rDNA spacer sequences		100,0	87,5
12	148	353 P$TEF1.01				
12	148	308 P$TEF1.01	TEF cis acting elements in both RNA polymerase II-dependent promoters and rDNA spacer sequences	148	88,6	78,4
12	148	434 P$TEF1.01	TEF cis acting elements in both RNA polymerase II-dependent promoters and rDNA spacer sequences	148	95,6	81,7
12	148	449 P$TEF1.01	TEF cis acting elements in both RNA polymerase II-dependent promoters and rDNA spacer sequences	148	95,6	85,7
12	149	321 P$TEIL.01	TEIL (tobacco EIN3-like)	149	96,4	92,6
12	149	378 P$TEIL.01	TEIL (tobacco EIN3-like)	149	96,4	92,6
12	149	427 P$TEIL.01	TEIL (tobacco EIN3-like)	149	100,0	92,1
12	150	325 P$TGA1.01	Arabidopsis leucine zipper protein TGA1	150	100,0	90,0
12	150	362 P$TGA1.01	Arabidopsis leucine zipper protein TGA1	150	100,0	90,2

12	150	437P$TGA1.01	Arabidopsis leucine zipper protein TGA1	100,0	90,0
12	153	334PSWRKY.01	WRKY plant specific zinc-finger-type factor associated with pathogen defence, W box	100,0	92,6
12	154	444PSWUS.01	Homeodomain protein WUSCHEL	100,0	96,3
12	155	328PSZAP1.01	Arabidopsis thaliana Zinc-dependent Activator Protein-1 (ZAP1)	100,0	88,0
13	26	202PSABF1.01	ABA (abscisic acid) inducible transcriptional activator	100,0	79,1
13	26	246PSABF1.01	ABA (abscisic acid) inducible transcriptional activator	100,0	81,2
13	30	242PSABRE.01	ABA response elements	100,0	87,1
13	34	168PSAGL16.01	AGL16, Arabidopsis MADS-domain protein AGAMOUS-like 16	100,0	83,0
13	34	189PSAGL16.01	AGL16, Arabidopsis MADS-domain protein AGAMOUS-like 16	92,5	82,5
13	34	191PSAGL16.01	AGL16, Arabidopsis MADS-domain protein AGAMOUS-	85,0	82,9

	AGL15, Arabidopsis MADS-domain protein AGAMOUS-like 15						100.0	84.9
12		196	PSAGL16.01	34				
	AGL15, Arabidopsis MADS-domain protein AGAMOUS-like 15						92.5	79.3
13		222	PSAGL15.01	34				
	AGL15, Arabidopsis MADS-domain protein AGAMOUS-like 15						85.0	81.9
13		226	PSAGL15.01	34				
	AGL15, Arabidopsis MADS-domain protein AGAMOUS-like 15						100.0	85.7
13		230	PSAGL16.01	34				
	AGL2, Arabidopsis MADS-domain protein AGAMOUS-like 2						96.8	95.4
13		229	PSAGL2.01	35				
	AOL3, MADS Box protein						100.0	84.9
13		107	PSAGL3.01	30				
	Zinc-finger protein in alfalfa roots, regulates salt tolerance						100.0	96.5
13		237	PSALFIN1.01	30				
	ANT (Arabidopsis protein AINTEGUMENTA), member of the plant-specific family of						100.0	83.7
13		161	PSANT.01	41				

AP2/EREBP-transcription factors						
						98.0
Motif involved in carotenoid and tocopherol biosynthesis and in the expression of photosynthesis-related genes					100.0	
	172PSATCTA.01					
		44		13		
						89.7
Motif involved in carotenoid and tocopherol biosynthesis and in the expression of photosynthesis-related genes					100.0	
	187PSATCTA.01					
		44		13		
						82.7
L1-specific homeodomain protein ATML1 (A. thaliana meristem layer 1)					75.0	
	192PSATML1.01					
		48		13		
						88.4
R2R3-type myb-like transcription factor (I-type binding site)					100.0	
	209PSATMYB77.01					
		50		13		
						86.8
Arabidopsis Telo-box interacting protein related to the conserved animal protein Pur-alpha					100.0	
	162PSATPURA.01					
		52		13		
						96.2
TCP class I transcription factor (Arabidopsis)					100.0	
	249PSATTCP20.01					
		54		13		

13	59	216 P$CAAT.01	CCAAT-box in plant promoters	100,0	98,5
13	61	168PSCARE.01	CAACTC regulatory elements, GA-inducible	100,0	84,3
13	61	221PSCARE.01	CAACTC regulatory elements, GA-inducible	100,0	83,0
13	61	258PSCARE.01	CAACTC regulatory elements, GA-inducible	100,0	88,0
13	61	259PSCARE.01	CAACTC regulatory elements, GA-inducible	100,0	87,0
13	62	188PSCCA1.01	Circadian clock associated 1	100,0	89,1
13	66	244PSCPRF.01	Common plant regulatory factor (CPRF) from parsley	100,0	99,0
13	66	245PSCPRF.01	Common plant regulatory factor (CPRF) from parsley	100,0	97,9
13	69	229PSDOF2.01	Dof2 - single zinc finger transcription factor	100,0	100,0
13	70	190PSDOF3.01	Dof3 - single zinc finger transcription factor	100,0	99,4
13	71	239PSDPBF.01	bZIP factors DPBF-1 and 2 (Dc3 promoter binding factor-1 and 2)	100,0	92,0
13	76	280PSERSE.I.01	ERSE I (ER stress-response element I)-like motif	75,0	82,8

			GAAA motif involved in pollen specific transcriptional activation	100,0	83,8
13	77	165 P$GAAA.01	GAAA motif involved in pollen specific transcriptional activation	100,0	86,2
13	77	177 PSGAAA.01	GAAA motif involved in pollen specific transcriptional activation	100,0	83,3
13	77	194 PSGAAA.01	GAAA motif involved in pollen specific transcriptional activation	100,0	95,6
13	77	189 PSGAAA.01	Cis-element in the GAPDH promoters conferring light inducibility	80,7	88,0
13	80	185 PSGAP.01	Cis-element in the GAPDH promoters conferring light inducibility	80,7	89,5
13	80	204 PSGAP.01	Cis-element in the GAPDH promoters conferring light inducibility	80,7	88,6
13	80	224 PSGAP.01	Cis-element in the GAPDH promoters conferring light	80,0	88,4
13	80	240 PSGAP.01			

13	81	174PSGATA.01	Class I GATA factors	100,0		96,3
13	81	184PSGATA.01	Class I GATA factors	100,0		94,9
13	81	217PSGATA.01	Class I GATA factors	100,0		93,2
13	81	230PSGATA.01	Class I GATA factors	100,0		94,3
13	82	241PSGBF1.01	bZIP protein G-Box binding factor 1	100,0		96,2
13	85	212PSGT1.01	GT1-Box binding factors with a trihelix DNA-binding domain	100,0		86,5
13	85	219PSGT1.01	GT1-Box binding factors with a trihelix DNA-binding domain	100,0		85,0
13	87	216PSHAHB4.01	Sunflower homeodomain leucine-zipper protein Hahb-4	100,0		90,8
13	87	237PSHAHB4.01	Sunflower homeodomain leucine-zipper protein Hahb-4	100,0		90,9
13	92	180PSHSE.01	Heat shock element	100,0		83,1
13	96	193PSICE.01	ICE (inducer of CBF expression 1), AtMYC2 (rd22BP1)	100,0		95,3
13	96	231PSICE.01	ICE (inducer of CBF expression 1), AtMYC2 (rd22BP1)	100,0		96,3

13	96	234 P$ICE.01	ICE (inducer of CBF expression 1), AtMYC2 (rd22BP1)	95,4	96,7
13	97	160 PSID1.01	Maize INDETERMINATE1 zinc finger protein	100,0	92,2
13	102	214 PSMSA.01	M-phase-specific activators (NtmybA1, NtmybA2, NtmybB)	100,0	89,6
13	102	227 PSMSA.01	M-phase-specific activators (NtmybA1, NtmybA2, NtmybB)	100,0	88,7
13	104	167 PSMYBPH3.01	Myb-like protein of Petunia hybrida	76,0	80,1
13	104	209 PSMYBPH3.01	Myb-like protein of Petunia hybrida	100,0	89,5
13	104	220 PSMYBPH3.01	Myb-like protein of Petunia hybrida	100,0	83,2
13	105	219 PSMYBPH3.02	Myb-like protein of Petunia hybrida	77,8	81,5
13	106	183 PSMYBST1.01	MybSt1 (Myb Solanum tuberosum 1) with a single myb repeat	100,0	95,4
13	106	215 PSMYBST1.01	MybSt1 (Myb Solanum tuberosum 1) with a single myb repeat	100,0	93,2

Nodulin consensus sequence	1				87,8	85,6
13	108	158	P$NCS1.01	Nodulin consensus sequence	100,0	95,2
13	108	161	PSNCS1.01	1		
13	106	165	PSNCS1.01	Nodulin consensus sequence	87,8	86,5
13	106	165	PSNCS1.01	1		
13	110	225	PSNCS3.01	Nodulin consensus sequence	100,0	89,7
13	110	225	PSNCS3.01	3		
13	114	229	PSO2_GCN4.01	Recognition site for BZIP transcription factors that belong to the group of Opaque-2 like proteins	100,0	81,2
13	116	178	PSOCSL01	OCS-like elements	76,9	71,0
13	119	229	PSOSMYBS.01	Rice MYB proteins with single DNA binding domains, binding to the amylase element (TATCCA)	100,0	91,8
13	121	234	PSPALBOXL.01	Cis-acting element conserved in various PAL and 4CL promoters	100,0	89,4
13	122	266	PSPALBOXP.01	Putative cis-acting element in various PAL and 4CL gene promoters	93,8	83,5

13	123	179	P$PBF.01	PBF (MPBF)	100,0	98,6
13	124	167	PSPBOX.01	Prolamin box, conserved in cereal seed storage protein gene promoters	77,6	79,6
13	124	175	PSPBOX.01	Prolamin box, conserved in cereal seed storage protein gene promoters	100,0	81,0
13	124	176	PSPBOX.01	Prolamin box, conserved in cereal seed storage protein gene promoters	76,1	75,0
13	124	246	PSPBOX.01	Prolamin box, conserved in cereal seed storage protein gene promoters	100,0	79,0
13	134	239	PSRY.01	RY and Sph motifs conserved in seed-specific promoters	100,0	88,2
13	134	247	PSRY.01	RY and Sph motifs conserved in seed-specific promoters	100,0	90,7
13	136	201	PSSBF1.01	SBF-1	100,0	87,3
13	136	208	PSSBF1.01	SBF-1	100,0	90,2
13	136	236	PSSBF1.01	SBF-1	100,0	89,7
13	138	253	PSSEF3.01	SEF3, Soybean embryo factor 3	100,0	93,7
13	138	285	PSSEF3.01	SEF3, Soybean embryo factor	100,0	93,3

3							
DNA-binding protein of sweet potato that binds to the SP8a (ACTGTCTA) and SP8b (TACTATT) sequences of sporamin and beta-amylase genes		184PSSP8BF.01	140	13	100.0	100.0	91.8
DNA-binding protein of sweet potato that binds to the SP8a (ACTGTCTA) and SP8b (TACTATT) sequences of sporamin and beta-amylase genes		182PSSP8BF.01	140	13	77.7	100.0	89.3
MADS-box protein SQUAMOSA		209PSSQUA.01	141	13	100.0	100.0	92.7
Storekeeper (STK) plant-specific DNA binding protein important for tuber-specific and sucrose-inducible gene expression		200PSSTK01	142	13	100.0	100.0	85.1
MYB protein from wheat		160PSTAMYB30.01	143	13	100.0	100.0	94.3
MYB protein from wheat		170PSTAMYB80.01	144	13	100.0	100.0	94.3

13	145	180	P$TANAC69.01	Wheat NACdomain DNA binding factor	81,2	69,4
13	146	210	PSTATA.01	Plant TATA box	100,0	95,4
13	146	252	PSTATA.01	Plant TATA box	100,0	92,2
13	147	180	PSTATA.02	Plant TATA box	100,0	98,8
13	147	202	PSTATA.02	Plant TATA box	100,0	90,2
13	147	206	PSTATA.02	Plant TATA box	100,0	92,3
13	147	211	PSTATA.02	Plant TATA box	100,0	91,2
13	147	251	PSTATA.02	Plant TATA box	100,0	95,4
13	146	207	PSTEF1.01	TEF cis acting elements in both RNA polymerase II-dependent promoters and rDNA spacer sequences	83,8	84,0
13	149	169	PSTEIL.01	TEIL (tobacco EIN3-like)	98,4	92,2
13	153	163	PSWRKY.01	WRKY plant specific zinc-finger-type factor associated with pathogen defence W box	100,0	94,3
13	154	171	PSWUS.01	Homeodomain protein WUSCHEL	100,0	100,0
13	154	172	PSWUS.01	Homeodomain protein WUSCHEL	100,0	100,0
13	150	249	PSZPT2.01	Member of the EPF family of zinc finger transcription factors	100,0	76,5

2063015119