

(12) STANDARD PATENT
(19) AUSTRALIAN PATENT OFFICE

(11) Application No. **AU 2010313865 B2**

- (54) Title
Modified oil encapsulating proteins and uses thereof
- (51) International Patent Classification(s)
C12N 15/29 (2006.01) **C12N 15/79** (2006.01)
C07K 14/415 (2006.01)
- (21) Application No: **2010313865** (22) Date of Filing: **2010.10.29**
- (87) WIPO No: **WO11/053169**
- (30) Priority Data
- (31) Number (32) Date (33) Country
61/256,689 **2009.10.30** **US**
- (43) Publication Date: **2011.05.05**
(44) Accepted Journal Date: **2016.05.05**
- (71) Applicant(s)
AgResearch Limited
- (72) Inventor(s)
Roberts, Nicholas John;Scott, Richard William;Winichayakul, Somrutai;Roldan, Marissa
- (74) Agent / Attorney
AJ PARK, L 9 Nishi 2 Phillip Law St, Canberra, ACT, 2601
- (56) Related Art
WO 2004/076673 A1 (BOREAL PLANT BREEDING LTD.) 10 September 2004

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
5 May 2011 (05.05.2011)

(10) International Publication Number
WO 2011/053169 A1

PCT

(51) International Patent Classification:
C12N 15/29 (2006.01) C12N 15/79 (2006.01)
C07K 14/415 (2006.01)

(74) Agents: ADAMS, Matthew, D et al.; A J Park, 6th Floor
Huddart Parker Building, Post Office Square, P O Box
949, Wellington, 6015 (NZ).

(21) International Application Number:
PCT/NZ2010/000218

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

(22) International Filing Date:
29 October 2010 (29.10.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/256,689 30 October 2009 (30.10.2009) US

(71) Applicant (for all designated States except US): AGRE-
SEARCH LIMITED [NZ/NZ]; 5th Floor, Tower Block,
Ruakura Research Centre, East Street, East Street Hamil-
ton (NZ).

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

(72) Inventors; and

(75) Inventors/Applicants (for US only): ROBERTS,
Nicholas, John [AU/NZ]; 97 McKay Line, Rd 7, Feild-
ing, 4777 (NZ). SCOTT, Richard, William [NZ/NZ]; 25
Colombo Street, Palmerston North, 4410 (NZ).
WINICHAYAKUL, Somrutai [TH/NZ]; 26 Geneva
Terrace, Kelvin Grove, Palmerston North, 4414 (NZ).
ROLDAN, Marissa [NZ/NZ]; 12 Brooklyn Heights
Drive, Kelvin Grove, Palmerston North, 4414 (NZ).

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: MODIFIED OIL ENCAPSULATING PROTEINS AND USES THEREOF

(57) Abstract: The invention provided modified oleosins including at least one artificially introduce cysteine. The invention provided methods and compositions for producing the modified oleosins. The invention provides polynucleotides encoding the modified oleosins, and constructs and host cells comprising the polynucleotides. The invention also provides methods for producing oil bodies, comprising the modified oleosins, in vivo and in vitro. The invention also provides methods for producing oil in host cells and plants. The invention also provides animal feed and biofuel sources comprising the oil bodies, host cells or plants of the invention.



WO 2011/053169 A1

MODIFIED NEUTRAL LIPID ENCAPSULATING PROTEINS AND USES THEREOF**TECHNICAL FIELD**

The invention relates to compositions and methods for the production and modification of oil bodies in various host cell types.

5 BACKGROUND

In nature, flowering plants efficiently store energy in their seeds through the accumulation of oil, namely triacylglycerol (TAG) and store it in discreet oil bodies by embedding a phospholipid protein monolayer around the oil body. These seed crops have been used in a variety of agricultural applications as feed and more recently also as a feedstock source for biofuels. On a
10 per weight basis, lipids have approximately double the energy content of either proteins or carbohydrates and as such, substantial focus has been placed on raising the oil content of various species, most notably plants. Beyond the energy aspect, the oil bodies themselves also have unique properties and form the basis for a number of biotechnical applications including but not limited to the purification of recombinant proteins, formation of multimeric protein complexes,
15 emulsification and the delivery of bio-actives.

Unfortunately plant seeds represent a very small percentage of total plant biomass and with the demand for improved agricultural productivity and alternative energies it is recognised that current oil production from a number of devoted seed crops is insufficient. Research efforts have focused on not only increasing the productivity of oil production within plant seeds but also oil
20 production in other cell types and species.

Traditional breeding and mutagenesis have offered incremental successes in this area; however genetic engineering has made the furthest strides in modifying organisms to produce elevated oil levels. While certain groups have worked along various parts of the oil synthesis pathway to up-regulate oil production within the seed, others groups have focused on increasing oil in cell types
25 that represent a larger portion of the biomass.

While genetic engineering has made some progress in increasing oil content in certain targets, significant challenges still remain. Further productivity increases can still be realized in oil body production in the seed and the means to produce oil bodies similar to those of a plant seed in other cell types and species has yet to be achieved.

SUMMARY OF THE INVENTION

The present invention provides compositions and methods for producing oil bodies with varying degrees of stability. The invention involves producing modified oleosins with artificially introduced cysteine residues. The artificially introduced cysteine residues are preferably

5 introduced in the N- and C-terminal hydrophilic arms of the modified oleosins.

Expression of the modified oleosins allows for the creation of stable oil bodies beyond the reproductive tissue of vascular plants into new cell types and even other species. When combined with a TAG synthesising enzyme, the invention leads to the accumulation and storage of TAG in eukaryotic cells as stable oil bodies. Compared with an unmodified cell or even one expressing
10 just a TAG synthesis enzyme, the invention allows for the accumulation of TAG in excess levels achieved by other means. For example the invention has shown that one can accumulate higher levels of stable oil bodies beyond the seed, in the vegetative portion of vascular plants.

Plants with increased levels of TAG in their vegetative tissues provide a valuable energy source for both animal feedstock and biofuel feedstock applications.

15 In addition recombinant modified oleosins purified from a host cell (such as *E. coli*, *P. pastoris*, *S. cerevisiae*, *Dunaliella*, *C. reinhardtii*) can be used to generate artificial oil bodies. The modified oleosins in artificial oil bodies, or those purified from transformed cells, can optionally be made to cross-link via the cysteine residues in the modified oleosin. The degree of cross-linking may be controlled manipulating the redox environment. The degree of cross-linking can
20 also be tailored by altering the number of cysteines in the modified oleosins.

Using combinations of these techniques the oil bodies formed with the modified oleosins can be tailored for their emulsification properties, to regulate thermal stability, chemical stability, and peptidase resistance.

The modified oleosins can also be fused to a protein of interest, to form a fusion protein. The
25 fusion protein (modified oleosin plus protein of interest) can be recombinantly expressed in a cell or organism. In this way oil bodies containing the expressed fusion proteins can be used to purify and deliver the protein of interest, for a variety of applications.

In addition the oil bodies can protect, or at least delay, degradation and/or biohydrogenation, of TAG, within the stomach and/or rumen of an animal, allowing the intact individual lipids from
30 the TAG to be absorbed by the animal in the intestine. Therefore, the invention is also useful in

terms of dietary intake of an animal, particularly through expression of the modified oleosins in plants.

Polynucleotides encoding modified oleosins with artificially introduced cysteines

In the first aspect the invention provides a polynucleotide encoding a modified oleosin including at least one artificially introduced cysteine, wherein the cysteine is introduced into at least one of a) the N-terminal hydrophilic region of the oleosin, and b) in the C-terminal hydrophilic region of the oleosin. The term oleosin also includes steroleosin and caloleosin. The modified oleosin may therefore be selected from a modified oleosin, a modified caloleosin or a modified steroleosin. In one embodiment the modified oleosin is a modified oleosin. In another embodiment the modified oleosin is a modified caloleosin. In another embodiment the modified oleosin is a modified steroleosin. Examples of each type of oleosin (oleosin, caloleosin and steroleosin) are described herein

In one embodiment, the modified oleosin includes at least two cysteines, at least one of which is artificially introduced. In a further embodiment, the modified oleosin includes at least two to at least thirteen (i.e., 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or more) artificially introduced cysteines. In one embodiment the cysteines are artificially introduced in the N-terminal hydrophilic region of the oleosin, or in the C-terminal hydrophilic region of the oleosin. In a further embodiment the modified oleosin includes at least one cysteine in the N-terminal hydrophilic region, and at least one cysteine in the C-terminal hydrophilic region. In a further embodiment the cysteines are distributed substantially evenly over the N-terminal and C-terminal hydrophilic regions of the oleosin.

In a further embodiment the polynucleotide encodes a fusion protein including the modified oleosin fused to a protein of interest.

Constructs

In a further aspect the invention provides a genetic construct comprising a polynucleotide of the invention. In a further aspect the invention provides an expression construct comprising a polynucleotide of the invention. In one embodiment the polynucleotide in the construct is operably linked to a promoter sequence. In one embodiment the promoter sequence is capable of driving expression of the polynucleotide in a vegetative tissue of a plant. In a further embodiment the promoter sequence is capable of driving expression of the polynucleotide in a seed of a plant. In a further embodiment the promoter sequence is capable of driving expression of the polynucleotide in the pollen of a plant. In a further embodiment the promoter sequence is capable of driving expression of the polynucleotide in an *E. coli* cell. In a further embodiment

the promoter sequence is capable of driving expression of the polynucleotide in a yeast cell. In a further embodiment the promoter sequence is capable of driving expression of the polynucleotide in an algal cell.

In another aspect, the invention provides a construct containing a polynucleotide that encodes a modified neutral lipid protein. In one embodiment, the construct also contains a second polynucleotide that encodes a triacylglycerol (TAG) synthesizing enzyme. In various embodiments, the construct can be linked to a promoter sequence capable of driving its expression in various host cells. As such, the invention also provides use of the constructs to induce a host cell to express a modified oleosin and/or a TAG synthesizing enzyme. In various embodiments, the construct expressing a modified oleosin and the construct expressing a TAG synthesizing enzyme may be driven by the same or by different promoters. In yet another embodiment the construct is located in an appropriate position and orientation of a suitable functional endogenous promoter such that the expression of the construct occurs. In various embodiments, the construct can be expressed in a bacterial, plant, fungal or algal cell. In one embodiment where the construct is expressed in a plant cell, the cell may be of vegetative, seed, pollen or fruit tissue.

Host cells

In a further aspect the invention provides a host cell comprising a construct of the invention. In a further aspect the invention provides a host cell genetically modified to comprise a polynucleotide of the invention. In a further aspect the invention provides a host cell genetically modified to express a polynucleotide of the invention.

Host cell also expressing a TAG synthesising enzyme

In a further embodiment the host cell is also genetically modified to express a triacylglycerol (TAG) synthesising enzyme. In a further embodiment the host cell is genetically modified to comprise a nucleic acid sequence encoding a triacylglycerol (TAG) synthesising enzyme. In a further embodiment the host cell comprises an expression construct including a nucleic acid sequence encoding a triacylglycerol (TAG) synthesising enzyme.

In a further embodiment the nucleic acid is operably linked to a promoter sequence. In a further embodiment the promoter sequence is capable of driving expression of the nucleic acid sequence in a vegetative tissue of a plant. In one embodiment the promoter sequence is capable of driving

expression of the nucleic acid sequence in a seed of a plant. In one embodiment the promoter sequence is capable of driving expression of the nucleic acid sequence in the pollen of a plant.

In a further embodiment the promoter sequence is capable of driving expression of the polynucleotide in an *E. coli* cell. In a further embodiment the promoter sequence is capable of driving expression of the polynucleotide in a yeast cell. In a further embodiment the promoter sequence is capable of driving expression of the polynucleotide in an algal cell.

Host cell types

The host cell may be any type of cell. In one embodiment the host cell is a prokaryotic cell. In a further embodiment the host cell is a eukaryotic cell. In one embodiment the host cell is selected from a bacterial cell, a yeast cell, a fungal cell, an insect cell, algal cell, and a plant cell. In one embodiment the host cell is a bacterial cell. In a further embodiment the host cell is a yeast cell. In further embodiment the host cell is a fungal cell. In further embodiment the host cell is an insect cell. In further embodiment the host cell is an algal cell. In a further embodiment the host cell is a plant cell.

Plants

In a further aspect the invention provides a plant comprising a plant cell of the invention. In a further aspect the invention provides a plant comprising a construct of the invention. In a further aspect the invention provides a plant genetically modified to comprise a polynucleotide of the invention. In a further aspect the invention provides a plant genetically modified to express a polynucleotide of the invention. In a further embodiment the plant expresses a modified oleosin encoded by the polynucleotide of the invention.

In a further embodiment the modified oleosin is expressed in a vegetative tissue of the plant. In a further embodiment the modified oleosin is expressed in a seed of the plant. In a further embodiment the modified oleosin is expressed in the pollen of the plant.

Plant also expresses a TAG enzyme

In a further embodiment the plant is also genetically modified to express a triacylglycerol (TAG) synthesising enzyme. In a further embodiment the triacylglycerol (TAG) synthesising enzyme is expressed in the same tissue as the modified oleosin.

In a further embodiment the plant is genetically modified to comprise a nucleic acid sequence encoding a triacylglycerol (TAG) synthesising enzyme. In a further embodiment the plant comprises an expression construct including a nucleic acid sequence encoding a triacylglycerol (TAG) synthesising enzyme.

In a further embodiment the nucleic acid is operably linked to a promoter sequence.

In a further embodiment the promoter sequence is capable of driving expression of the nucleic acid sequence in a vegetative tissue of a plant. In one embodiment the promoter sequence is capable of driving expression of the nucleic acid sequence in a seed of a plant. In one embodiment the promoter sequence is capable of driving expression of the nucleic acid sequence in the pollen of a plant.

Modified oleosin polypeptides with artificially introduced cysteines

In a further aspect the invention provides a modified oleosin including at least one artificially introduced cysteine, wherein the cysteine is introduced into at least one of a) the N-terminal hydrophilic region of the oleosin, and b) in the C-terminal hydrophilic region of the oleosin. In a further aspect the invention provides a modified oleosin encoded by a polynucleotide of the invention. In one embodiment, the modified oleosin includes at least two cysteines, at least one of which is artificially introduced. In a further embodiment, the modified oleosin includes at least two to at least thirteen (i.e., 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or more) artificially introduced cysteines.

In a further embodiment the modified oleosin includes at least one cysteine in the N-terminal hydrophilic region, and at least one cysteine in the C-terminal hydrophilic region. In a preferred embodiment the cysteines are artificially introduced in the N-terminal hydrophilic region of the oleosin, or in the C-terminal hydrophilic region of the oleosin. Preferably the cysteines are distributed substantially evenly between the N-terminal and C-terminal hydrophilic region of the oleosin.

Fusion proteins with modified oleosins including artificially introduced cysteines

In a further aspect the invention provides a fusion protein comprising a modified oleosin of the invention and a protein of interest. The fusion protein thus comprises a modified oleosin portion, and a protein of interest portion.

Oil bodies comprising modified oleosins

In a further aspect the invention provides an oil body comprising a modified oleosin of the invention. In a further aspect the invention provides an oil body comprising at least two modified oleosins of the invention. In one embodiment at least two of the modified oleosins are cross-linked to each other via disulphide bridges between cysteine residues in the modified
5 oleosins. In a further embodiment the modified oleosins are cross-linked via the artificially introduced cysteine residues in the modified oleosins.

In a further embodiment the oil body additionally comprises a fusion protein, wherein the fusion protein includes an oleosin fused to a protein of interest. In this embodiment, the oleosin in the
10 fusion protein need not include an artificially introduced cysteine. Preferably the oleosin in the fusion protein does not include an artificially introduced cysteine.

The oil bodies of this embodiment are useful for purifying and delivering the protein of interest, as discussed in Roberts *et al.*, (2008).

However in this embodiment it is possible to take advantage of the option to vary the
15 stability/integrity of the oil body provided by presence of the modified oleosins in the oil body, hence allowing for more stringent purification and delivery procedures.

Oil bodies comprising fusion proteins with modified oleosins

In a further aspect the invention provides an oil body comprising a fusion protein of the invention, the fusion protein comprising a modified oleosin of the invention and a protein of
20 interest. The fusion protein thus comprises a modified oleosin portion, and a protein of interest portion.

In one embodiment the oil body comprises at least two fusion proteins of the invention.

In one embodiment at least two of the fusion proteins are cross-linked to each other via disulphide bridges between cysteine residues in the modified oleosin portion of the fusion
25 proteins. In one embodiment the fusion proteins are cross-linked via the artificially introduced cysteine residues in the modified oleosin portion of the fusion proteins.

In a further embodiment the oil body comprises at least one modified oleosin of the invention. In a further embodiment at least one fusion protein is cross-linked to at least one modified oleosin,

via a cysteine in the modified oleosin portion of the fusion protein and a cysteine in the modified oleosin.

Again, the oil bodies of this embodiment are useful for purifying and delivering the protein of interest, as discussed in Roberts *et al.*, (2008).

- 5 However in this embodiment it is possible to take advantage of the option to vary the stability/integrity of the oil body provided by presence of the modified oleosins in the oil body, hence allowing for more stringent purification and delivery procedures.

Emulsion

- 10 In a further aspect the invention provides an emulsion comprising a modified oleosin of the invention. In one embodiment the emulsion comprises the modified oleosin and a suitable carrier. The carrier may be buffered, with the appropriate redox environment to retain the desired degree of cross-linking of the oleosins.

To resuspend the modified oleosin in the carrier may require sonication or high pressure homogenising, followed by exposure to the appropriate oxidising conditions.

15 *Compositions*

In a further aspect the invention provides a composition comprising a modified oleosin of the invention. In one embodiment the composition comprises the modified oleosin and a suitable carrier. The carrier may be buffered, with the appropriate redox environment to attain the desired degree of cross-linking of the modified oleosins.

- 20 To resuspend the modified oleosins in the carrier may require sonication or high pressure homogenising, followed by exposure to the appropriate oxidising conditions.

- In a further aspect the invention provides a composition comprising an oil body of the invention. In one embodiment the composition comprises the oil body and a suitable carrier. The carrier may be buffered, with the appropriate redox environment to retain the desired degree of cross-linking of the modified oleosins. In a further embodiment the invention provides a composition
25 formulated for dermal application comprising an oil body of the invention.

Plants, and parts thereof, comprising oil bodies of the invention

In a further aspect the invention provides a plant, or part thereof, comprising an oil body of the invention. In a further aspect the invention provides a vegetative tissue of a plant, comprising an oil body of the invention. In a further aspect the invention provides a seed of a plant, comprising an oil body of the invention.

Animal feed comprising oil bodies of the invention

In a further aspect the invention provides an animal feed comprising an oil body of the invention. In a further aspect the invention provides an animal feed comprising a plant, or part thereof, of the invention.

10 *Methods for producing oil bodies*

In a further aspect invention provides a method for producing an oil body, the method comprising the step of combining:

- a) at least two modified oleosins, each including at least one artificially introduced cysteine,
- b) triacylglycerol, and
- 15 c) phospholipid.

In one embodiment, the modified oleosins each include at least two cysteines, at least one of which is artificially introduced. In a further embodiment the modified oleosins each include at least one cysteine in the N-terminal hydrophilic region of the oleosin, and at least one cysteine in the C-terminal hydrophilic region of the oleosin.

20 In a further embodiment, the modified oleosin includes at least two to at least thirteen (i.e., 2, 3, 4, 5, 6, 7, 8, 9, 10, 1, 12, 13, 14 or more) artificially introduced cysteines.

In one embodiment the cysteines are artificially introduced in the N-terminal hydrophilic region of the oleosins, or in the C-terminal hydrophilic region of the oleosins. In a further embodiment the cysteines are distributed substantially evenly between the N-terminal and C-terminal hydrophilic region of the oleosins. In a further embodiment the modified oleosins are cross-linked via disulphide bridges between cysteine residues in the oleosins. In a further embodiment 25 embodiment the modified oleosins are cross-linked between the artificially introduced cysteine residues in the oleosins.

In one embodiment the modified oleosins are part of fusion proteins wherein the fusion proteins comprise a modified oleosin, and a protein of interest.

In one embodiment the method comprises the additional step of regulating the degree of cross-linking of modified oleosins in the oil body by controlling the redox environment of the oil body
5 produced.

All components combined in vivo (in vivo oil bodies)

In one embodiment the components of a), b) and c) are combined within a host cell. In this embodiment the modified oleosins are preferably expressed in the host cell.

The host cell is preferably genetically modified to express the modified oleosins.

10 The host cell is preferably comprises a construct of the invention. The host cell is preferably genetically modified to comprise a polynucleotide of the invention. The host cell is preferably genetically modified to express a polynucleotide of the invention.

Host cell also expresses a TAG synthesising enzyme

In a further embodiment the host cell is also genetically modified to express a triacylglycerol
15 (TAG) synthesising enzyme. In a further embodiment the host cell comprises an expression construct including a nucleic acid sequence encoding a triacylglycerol (TAG) synthesising enzyme.

In a further embodiment the nucleic acid sequence is operably linked to a promoter sequence. In one embodiment the promoter sequence is capable of driving expression of the nucleic acid
20 sequence in a vegetative tissue of a plant. In one embodiment the promoter sequence is capable of driving expression of the nucleic acid sequence in a seed of a plant. In one embodiment the promoter sequence is capable of driving expression of the nucleic acid sequence in the pollen of a plant.

In a further embodiment the host cell is also genetically modified to comprise a nucleic acid
25 sequence encoding a triacylglycerol (TAG) synthesising enzyme. In a further embodiment the host cell is also genetically modified to express a nucleic acid sequence encoding a triacylglycerol (TAG) synthesising enzyme.

It will be understood by those skilled in the art that the polynucleotide encoding the modified oleosin and the nucleic acid sequence encoding a triacylglycerol (TAG) synthesising enzyme can

be placed on the same construct or on separate constructs to be transformed into the host cell. Expression of each can be driven by the same or different promoters, which may be included in the construct to be transformed. It will also be understood by those skilled in the art that alternatively the polynucleotide and nucleic acid can be transformed into the cell without a promoter, but expression of either the polynucleotide and nucleic acid could be driven by a promoter or promoters endogenous to the cell transformed.

In a further embodiment the host cell forms part of an organism. In a preferred embodiment the organism is a plant.

In a further embodiment the oil is produced in the vegetative tissues of the plant.

In one embodiment of the method the plant acumulates about 50% to about 400% more lipid than does a suitable control plant. In a further embodiment of the method the plant acumulates about 100% to about 300% more lipid than does a suitable control plant. In a further embodiment of the method the plant acumulates about 150% to about 250% more lipid than does a suitable control plant. Suitable control plants include non-transformed or wild-type versions of plant of the same variety and or species as the transformed plant used in the method of the invention.

In a further embodiment the plant is processed into an animal feed.

In a further embodiment the plant is processed into a biofuel feed stock.

Additional method step to purify the in vivo produced oil bodies

In one embodiment the method includes the additional step of purifying the oil bodies from the cell or organism.

Additional method step to vary degree of cross-linking of in vivo produced purified oil bodies

In a further embodiment the method comprises the additional step of regulating the degree of cross-linking of modified oleosins in the *in vivo* produced purified oil bodies by controlling the redox environment of the purified oil bodies. In one embodiment the degree of cross-linking is increased by use of an oxidising environment. In a further embodiment the degree of cross-linking is decreased by use of a reducing environment.

Components combined in vitro (in vitro / artificial oil bodies)

In certain embodiments the components of a), b) and c) may be combined *in vitro*.

In one embodiment, the modified oleosin of a) has been recombinantly expressed in, and purified from a host cell of the invention, before being combined with the components of b) and c).

5 *Additional method step to vary degree of cross-linking of in vitro / artificial oil bodies*

In a further embodiment the method comprises the additional step of regulating the degree of cross-linking by controlling the redox environment in which the components of a), b) and c) are combined. In one embodiment the degree of cross-linking is increased by combining the components of a), b) and c) in an oxidising environment. In a further embodiment the degree of
10 cross-linking is decreased by combining the components of a), b) and c) in a reducing environment. The degree of cross-linking may also be regulated after the oil body is formed, by controlling the redox environment in which the oil body is contained.

In a further aspect the invention provides a method of producing a plant that accumulates more oil than a suitable control plant the method comprising providing a plant transformed with a
15 polynucleotide of the invention that expresses a modified oleosin encoded by the polynucleotide.

In one embodiment the plant is also transformed with a polynucleotide encoding a TAG synthesising enzyme to express the TAG synthesising enzyme and thus synthesise TAG.

In one embodiment the plant the plant is produced by transforming a single plant, or plant cell, with both the polynucleotide of any one of the invention and the polynucleotide encoding the TAG
20 synthesising enzyme.

In a further embodiment the plant is produced by crossing a first plant transformed with a polynucleotide of any one of the invention, with a second plant transformed with the polynucleotide encoding the TAG synthesising enzyme, to produce the plant transformed with both a polynucleotide of the invention, and a polynucleotide encoding the TAG synthesising enzyme.

25 In a further embodiment the oil is TAG. In a further embodiment the oil is produced in the vegetative tissues of the plant.

In one embodiment of the method the plant accumulates about 50% to about 400% more lipid than does a suitable control plant. In a further embodiment of the method the plant accumulates about 100% to about 300% more lipid than does a suitable control plant. In a further

embodiment of the method the plant accumulates about 150% to about 250% more lipid than does a suitable control plant

In a further embodiment the plant is processed into an animal feed.

In a further embodiment the plant is processed into a biofuel feed stock.

- 5 In a further aspect invention provides a method for producing an oil body in a host cell, the method comprising:
- a) introducing into a host cell at least one nucleic acid molecule encoding a modified oleosin of the invention; and
 - b) culturing the host cell in order to express the modified oleosin.
- 10 In a further aspect invention provides a method for producing an oil body in a host cell, the method comprising:
- a) introducing into a host cell at least one nucleic acid molecule encoding a modified oleosin of the invention and a nucleic acid molecule encoding a TAG synthesizing enzyme ; and
 - 15 b) culturing the host cell in order to express the modified oleosin and the TAG synthesizing enzyme.

The host cell may be a host cell as herein described.

Oil bodies

In a further aspect invention provides an oil body produced by a method of the invention.

20 *Compositions*

In a further aspect the invention provides a composition comprising an oil body of the invention. In one embodiment the composition comprises the oil body and a suitable carrier. The carrier may be buffered to provide the appropriate redox environment to retain the desired degree of cross-linking of the modified oleosin. In a further embodiment the invention provides a
25 composition formulated for dermal application comprising an oil body of the invention.

Plants, and parts thereof, comprising oil bodies of the invention

In a further aspect the invention provides a plant, or part thereof, comprising an oil body of the invention. In a further aspect the invention provides a vegetative tissue of a plant, comprising an

oil body of the invention. In a further aspect the invention provides a seed of a plant, comprising an oil body of the invention. In a further aspect the invention provides pollen of a plant, comprising an oil body of the invention. In a further aspect the invention provides a fruit, or fruiting body, of a plant, comprising an oil body of the invention.

5 *Animal feed comprising oil bodies of the invention*

In a further aspect the invention provides an animal feed comprising an oil body of the invention. In a further aspect the invention provides an animal feed comprising a plant, or part thereof, of the invention.

10 In one embodiment the feed is suitable for a mammalian animal including humans. In a further embodiment the feed is suitable for non-human mammals. Preferred animals include farm animals such as but not limited to cows, sheep, horses, goats, pigs, chickens, and the like.

Plants

15 The modified oleosins may be modified naturally occurring oleosins. The plants from which the un-modified oleosin sequences are derived may be from any plant species that contains oleosins and polynucleotide sequences encoding oleosins.

The plant cells, in which the modified oleosins are expressed, may be from any plant species. The plants, in which the modified oleosins are expressed, may be from any plant species.

In one embodiment the plant cell or plant, is derived from a gymnosperm plant species.

In a further embodiment the plant cell or plant, is derived from an angiosperm plant species.

20 In a further embodiment the plant cell or plant, is derived from a from dicotyledonous plant species.

In a further embodiment the plant cell or plant, is derived from a monocotyledonous plant species.

25 Other preferred plants are forage plant species from a group comprising but not limited to the following genera: *Zea*, *Lolium*, *Hordium*, *Miscanthus*, *Saccharum*, *Festuca*, *Dactylis*, *Bromus*, *Thinopyrum*, *Trifolium*, *Medicago*, *Pheleum*, *Phalaris*, *Holcus*, *Glycine*, *Lotus*, *Plantago* and *Cichorium*.

Other preferred plants are leguminous plants. The leguminous plant or part thereof may encompass any plant in the plant family Leguminosae or Fabaceae. For example, the plants may be selected from forage legumes including, alfalfa, clover; leucaena; grain legumes including, beans, lentils, lupins, peas, peanuts, soy bean; bloom legumes including lupin, pharmaceutical or industrial legumes; and fallow or green manure legume species.

A particularly preferred genus is *Trifolium*. Preferred *Trifolium* species include *Trifolium repens*; *Trifolium arvense*; *Trifolium affine*; and *Trifolium occidentale*. A particularly preferred *Trifolium* species is *Trifolium repens*.

Another preferred genus is *Medicago*. Preferred *Medicago* species include *Medicago sativa* and *Medicago truncatula*. A particularly preferred *Medicago* species is *Medicago sativa*, commonly known as alfalfa.

Another preferred genus is *Glycine*. Preferred *Glycine* species include *Glycine max* and *Glycine wightii* (also known as *Neonotonia wightii*). A particularly preferred *Glycine* species is *Glycine max*, commonly known as soy bean. A particularly preferred *Glycine* species is *Glycine wightii*, commonly known as perennial soybean.

Another preferred genus is *Vigna*. A particularly preferred *Vigna* species is *Vigna unguiculata* commonly known as cowpea.

Another preferred genus is *Mucana*. Preferred *Mucana* species include *Mucana pruriens*. A particularly preferred *Mucana* species is *Mucana pruriens* commonly known as velvetbean.

Another preferred genus is *Arachis*. A particularly preferred *Arachis* species is *Arachis glabrata* commonly known as perennial peanut.

Another preferred genus is *Pisum*. A preferred *Pisum* species is *Pisum sativum* commonly known as pea.

Another preferred genus is *Lotus*. Preferred *Lotus* species include *Lotus corniculatus*, *Lotus pedunculatus*, *Lotus glabar*, *Lotus tenuis* and *Lotus uliginosus*. A preferred *Lotus* species is *Lotus corniculatus* commonly known as Birdsfoot Trefoil. Another preferred *Lotus* species is *Lotus glabar* commonly known as Narrow-leaf Birdsfoot Trefoil. Another preferred *Lotus* species is *Lotus pedunculatus* commonly known as Big trefoil. Another preferred *Lotus* species is *Lotus tenuis* commonly known as Slender trefoil.

Another preferred genus is *Brassica*. A preferred *Brassica* species is *Brassica oleracea*, commonly known as forage kale and cabbage.

Other preferred species are oil seed crops including but not limited to the following genera: *Brassica*, *Carthamus*, *Helianthus*, *Zea* and *Sesamum*.

5 A preferred oil seed genera is *Brassica*. A preferred oil seed species is *Brassica napus*.

A preferred oil seed genera is *Brassica*. A preferred oil seed species is *Brassica oleraceae*.

A preferred oil seed genera is *Zea*. A preferred oil seed species is *Zea mays*.

A preferred oil seed genera is *Carthamus*. A preferred oil seed species is *Carthamus tinctorius*.

A preferred oil seed genera is *Helianthus*. A preferred oil seed species is *Helianthus annuus*.

10 A preferred oil seed genera is *Zea*. A preferred oil seed species is *Zea mays*.

A preferred oil seed genera is *Sesamum*. A preferred oil seed species is *Sesamum indicum*.

A preferred silage genera is *Zea*. A preferred silage species is *Zea mays*.

A preferred grain producing genera is *Hordeum*. A preferred grain producing species is *Hordeum vulgare*.

15 A preferred grazing genera is *Lolium*. A preferred grazing species is *Lolium perenne*.

A preferred grazing genera is *Lolium*. A preferred grazing species is *Lolium arundinaceum*.

A preferred grazing genera is *Trifolium*. A preferred grazing species is *Trifolium repens*.

A preferred grazing genera is *Hordeum*. A preferred grazing species is *Hordeum vulgare*.

20 Preferred plants also include forage, or animal feedstock plants. Such plants include but are not limited to the following genera: *Miscanthus*, *Saccharum*, *Panicum*.

A preferred biofuel genera is *Miscanthus*. A preferred biofuel species is *Miscanthus giganteus*.

A preferred biofuel genera is *Saccharum*. A preferred biofuel species is *Saccharum officinarum*.

A preferred biofuel genera is *Panicum*. A preferred biofuel species is *Panicum virgatum*.

DETAILED DESCRIPTION OF THE INVENTION

In this specification where reference has been made to patent specifications, other external documents, or other sources of information, this is generally for the purpose of providing a context for discussing the features of the invention. Unless specifically stated otherwise, reference to such external documents is not to be construed as an admission that such documents, or such sources of information, in any jurisdiction, are prior art, or form part of the common general knowledge in the art.

The term "comprising" as used in this specification means "consisting at least in part of". When interpreting each statement in this specification that includes the term "comprising", features other than that or those prefaced by the term may also be present. Related terms such as "comprise" and "comprises" are to be interpreted in the same manner.

On a weight for weight basis lipids have approximately double the energy content of either proteins or carbohydrates. The bulk of the world's lipids are produced by plants and the densest form of lipid is as a triacylglycerol (TAG). Dicotyledonous plants can accumulate up to approximately 60% of their seed weight as TAG which is subsequently used as an energy source for germination. As such there have been a number of efforts targeted at using seeds rich in oils to sustainably produce sufficient lipids for both animal and biofuel feed stock.

Given that there is only a limited quantity of TAG able to be produced by seeds alternative approaches are being made to produce additional lipid (preferentially TAGs) in vegetative tissues. The majority of these approaches have pursued the up regulation or over expression of one or several enzymes in the Kennedy pathway in the leaves of plants in order to synthesise TAG. Typically however, the majority of additional lipids produced by this approach are remobilised within the plant by a combination of lipases and β -oxidation resulting in a limited increase in lipid content (usually 2-4% of the DM).

The TAG produced in developing seeds is typically contained within discrete structures called oil bodies (OBs) which are highly stable and remain as discrete tightly packed organelles without coalescing even when the cells desiccate or undergo freezing conditions (Siloto et al., 2006; Shimada et al., 2008). OBs consist of a TAG core surrounded by a phospholipid monolayer embedded with proteinaceous emulsifiers. The latter make up 0.5-3.5% of the OB; of this, 80-90% is oleosin with the remainder predominantly consisting of the calcium binding (caloleosin) and sterol binding (steroleosin) proteins (Lin and Tzen, 2004). The emulsification properties of oleosins derives from their three functional domains which consist of an amphipathic N-terminal

arm, a highly conserved central hydrophobic core (~72 residues) and a C-terminal amphipathic arm. Similarly, both caleosin and steroleosin possess hydrophilic N and C-terminal arms and their own conserved hydrophobic core.

It was previously speculated that the constitutive expression of oleosin or polyoleosin (tandem head-to-tail fusions of oleosins) with TAG synthesising enzymes in the leaves would result in the formation of stable oil bodies leading to the accumulation of TAG. We have subsequently found however, that oleosin and polyoleosins are ineffective and promoting the accumulation of TAG when co-expressed with DGAT1 in plant leaves (Roberts *et al.*, unpublished data).

The current invention provides modified oleosins which contain one or more artificially introduced cysteine residues. The encapsulation of the neutral lipids by oleosins containing engineered cysteines provides an alternative mechanism to accumulate appreciable quantities of TAG in leaves without the requirement to wait until senescence and without producing extreme phenotypes. In addition the modified oleosin has a number of other applications involving modifying OB stability, emulsion properties as well as the generation and purification of recombinant proteins.

Oil bodies

OBs generally range from 0.5-2.5µm in diameter and consist of a TAG core surrounded by a phospholipid monolayer embedded with proteinaceous emulsifiers - predominantly oleosins (Tzen et al, 1993; Tzen, et al 1997). OBs consist of only 0.5-3.5% protein; of this 80-90% is oleosin with the remainder predominantly consisting of the calcium binding (caleosin) and sterol binding (steroleosin) proteins (Lin and Tzen, 2004). The ratio of oleosin to TAG within the plant cell influences the size and number of oil bodies within the cell (Sarmiento et al., 1997; Siloto et al., 2006).

While OBs are naturally produced predominantly in the seeds and pollen of many plants they are also found in some other organs (e.g., specific tubers).

Oleosins are comparatively small (15-24 kDa) proteins that are embedded in the surface of OBs.

Oil body stability

The suitability of oil bodies, and artificial oil bodies, for the applications discussed above, among others, is limited at least in part, by their stability. One approach to address oil body stability was to generate oil bodies comprising so-called polyoleosin. Polyoleosin is the head to tail fusion of

two or more oleosin units (Roberts et al., 2008). Altering the number of oleosin units enables the properties (thermal stability and degradation rate) of the oil bodies to be tailored. Expression of polyoleosin *in planta* leads to incorporation of the polyoleosin units to the oil bodies as per single oleosin units (Scott et al., 2007). Multiple oleosin units in tandem head-to-tail arrangements
5 were used to create polyoleosin. Separate constructs (containing from one to six oleosin repeats) were specifically designed for expression *in planta* and in *E. coli*. The majority of recombinant polyoleosin accumulated in the oil bodies of transgenic plants and in the inclusion bodies of *E. coli*. Purified prokaryotically produced polyoleosin was used to generate artificial oil bodies. Oil body and artificial oil body thermal stability and structural integrity in proteinase-K were raised
10 by polyoleosin.

However, there are several limiting factors determining the degree of protection/stability that polyoleosin can provide; these relate to the number of tandem repeats that can be joined before the process of translation and oil body targeting becomes limiting (Scott et al., 2007); while another limitation comes from the nature of the oleosin fusion which is achieved by generating a
15 transcript with a head to tail fusion arrangement. This is essentially a linear protein of multimeric oleosin repeats that has a number of covalent-links and position of covalent-links per individual oleosin repeat (i.e., a maximum of one at each end). In addition this arrangement only affords protection against N-terminal degrading proteins but it does not provide any additional protection against other proteolytic enzymes that recognise specific internal peptide sequences.
20 Furthermore, the linking between oleosin units in a polyoleosin molecule formed by tandem head to tail repeats is not readily altered *in situ*. While specific protease specific sites could be engineered into the joining regions in order to break apart fused polyoleosin molecules embedded into an oil body or artificial oil body they could not be re-fused easily.

Oleosins embedded in oil bodies have previously covalently cross-linked by the addition of
25 cross-linking agents such as glutaraldehyde or genepin (Peng et al., 2004 & 2006), however, this random cross-linking requires the addition of cross-linking agents to oil body preparations, and is not easy to reverse.

Artificial oil bodies

Prokaryotically expressed recombinant oleosins can be used to generate artificial oil bodies
30 (AOBs) whose properties are very similar to plant derived OBs (Peng et al. 2004; Roux et al. 2004; Chiang et al. 2005; Chiang et al. 2007).

Applications of oil bodies and artificial oil bodies

The unique properties of oil bodies, and their constituent oleosins, form the basis of a number of biotechnical applications including: purifying recombinant proteins; formation of multimeric protein complexes; emulsification; delivery of bioactives; generation of multivalent bioactives and even as a potential flavour enhancer (for reviews see Capuano et al., 2007 and Roberts et al., 2008).

Emulsions

Emulsions are produced when one or more liquids that are immiscible in another liquid, usually due to different polarities and thus different hydrophobicities, are uniformly suspended within that liquid. Examples include oil droplets uniformly dispersed in water, or water droplets uniformly dispersed in oil. Generation of a relatively stable emulsion requires the use of an emulsifier, which lowers the interfacial tension between the liquids. The stability of an emulsion is generally measured in terms of the duration that the uniform dispersion persists under specified conditions. Emulsifiers are commonly used in the food and cosmetic industry; so need to have high emulsion stability and be safe for consumption and topical application.

Intact oil bodies containing oleosin naturally form a surfactant-free, oil-in water emulsion. It has been found that intact oil bodies or oil bodies in which the majority of TAG has been removed have a broad range of emulsification applications in food, topical personal care (skin creams) and pharmaceutical formulations (Harada et al., 2002; Deckers et al., 2003; Hou et al., 2003).

Biohydrogenation

It has been demonstrated that the lipid profile of ruminant animal feed in turn influences the lipid profile of meat and dairy products (Demeyer and Doreau, 1999). Different plants have different lipid profiles; by selectively feeding animals only plants with the desired lipid profile it is possible to positively influence the lipid profile of downstream meat and dairy products. In ruminants the final lipid make up of the meat and milk is not only influenced by the dietary lipids but is also heavily influenced by biohydrogenation (Jenkins and McGuire 2006; Firkins et al., 2006; Lock and Bauman, 2004). Biohydrogenation is the hydrogenation of non-reduced compounds (such as unsaturated fats) by the biota present in the rumen. Biohydrogenation can be prevented/delayed by encapsulating the lipids in a protein or proteins that provide resistance to microbial degradation (Jenkins and Bridges 2007). The prevention of biohydrogenation by

encapsulating triacylglycerides in polyoleosin or oleosins *in planta* was reported by Scott et al., (2007), Cookson et al., (2009) and Roberts et al., (2008).

Oleosins

Oleosins are comparatively small (15 to 24 kDa) proteins which allow the OBs to become tightly packed discrete organelles without coalescing as the cells desiccate or undergo freezing conditions (Leprince *et al.*, 1998; Siloto *et al.*, 2006; Slack *et al.*, 1980; Shimada *et al.* 2008).

Oleosins have three functional domains consisting of an amphipathic N-terminal arm, a highly conserved central hydrophobic core (~72 residues) and a C-terminal amphipathic arm. The accepted topological model is one in which the N- and C-terminal amphipathic arms are located on the outside of the OBs and the central hydrophobic core is located inside the OB (Huang, 1992; Loer and Herman, 1993; Murphy 1993). The negatively charged residues of the N- and C-terminal amphipathic arms are exposed to the aqueous exterior whereas the positively charged residues are exposed to the OB interior and face the negatively charged lipids. Thus, the amphipathic arms with their outward facing negative charge are responsible for maintaining the OBs as individual entities via steric hinderance and electrostatic repulsion both *in vivo* and in isolated preparation (Tzen et al, 1992). The N-terminal amphipathic arm is highly variable and as such no specific secondary structure can describe all examples. In comparison the C-terminal arm contains a α -helical domain of 30-40 residues (Tzen et al, 2003). The central core is highly conserved and thought to be the longest hydrophobic region known to occur in nature; at the center is a conserved 12 residue proline knot motif which includes three spaced proline residues (for reviews see Frandsen et al, 2001; Tzen et al, 2003). The secondary, tertiary and quaternary structure of the central domain is still unclear. Modelling, Fourier Transformation-Infra Red (FT-IR) and Circular Dichromism (CD) evidence exists for a number of different arrangements (for review see Roberts et al., 2008).

The properties of the major oleosins is relatively conserved between plants and is characterised by the following:

- 15-25kDa protein corresponding to approximately 140-230 amino acid residues.
- The protein sequence can be divided almost equally along its length into 4 parts which correspond to a N-terminal hydrophilic region, two centre hydrophobic regions (joined by a proline knot or knob) and a C-terminal hydrophilic region.
- The topology of oleosin is attributed to its physical properties which includes a folded hydrophobic core flanked by hydrophilic domains. This arrangement confers an

amphipathic nature to oleosin resulting in the hydrophobic domain being embedded in the phospholipid monolayer (Tzen et al., 1992) while the flanking hydrophilic domains are exposed to the aqueous environment of the cytoplasm.

- Typically oleosins do not contain cysteines

- 5 Preferred oleosins for use in the invention are those which contain a central domain of approximately 70 non-polar amino acid residues (including a proline knot) uninterrupted by any charged residues, flanked by two hydrophilic arms.

The term "oleosin" as used herein also includes steroleosin and caloleosin

Steroleosins

- 10 Steroleosins comprises an N-terminal anchoring segment comprising two amphipathic α -helices (912 residues in each helix) connected by a hydrophobic anchoring region of 14 residues. The soluble dehydrogenase domain contains a NADP⁺- binding subdomain and a sterol-binding subdomain. The apparent distinction between steroleosins-A and -B occurs in their diverse sterol-binding subdomains (Lin and Tzen, 2004). Steroleosins have a proline knob in their
- 15 hydrophobic domain and contains a sterol-binding dehydrogenase in one of their hydrophilic arms.

Caloleosins

- Caloleosins (Frandsen *et al.*, 2001) have a slightly different proline knot than do the basic oleosins, and contain a calcium-binding motif and several potential phosphorylation sites in the
- 20 hydrophilic arms. Similar to oleosin, caloleosin is proposed to have three structural domains, where the N- and C-terminal arms are hydrophilic while the central domain is hydrophobic and acts as the oil body anchor. The N-terminal hydrophilic domain consists of a helix-turn-helix calcium binding EF-hand motif of 28 residues including an invariable glycine residue as a structural turning point and five conserved oxygen-containing residues as calcium-binding
- 25 ligands (Chen *et al.*, 1999; Frandsen *et al.*, 2001). The C-terminal hydrophilic domain contains several phosphorylation sites and near the C-terminus is an invariable cysteine that is not involved in any intra- or inter-disulfide linkages (Peng, 2004). The hydrophilic N- and C-termini of caloleosin are approximately 3 times larger than those of oleosin (Lin and Tzen, 2004). The hydrophobic domain is thought to consist of an amphipathic α -helix and an anchoring region
- 30 (which includes a proline knot).

Examples of oleosin (oleosins, steroleosin and caloleosin) sequences suitable to be modified for use in the invention, by the addition of at least one artificially introduced cysteine, are shown in Table 1 below. The sequences (both polynucleotide and polypeptide are provided in the Sequence Listing)

5 **Table 1**

Oleosin	Species	cDNA accession no.	SEQ ID NO:	Protein accession no.	SEQ ID NO:
Oleosin	<i>S. indicum</i>	AF302907	34	AAG23840	35
Oleosin	<i>S. indicum</i>	U97700	36	AAB58402	37
Oleosin	<i>A. thaliana</i>	X62353	38	CAA44225	39
Oleosin	<i>A. thaliana</i>	BT023738	40	AAZ23930	41
Oleosin	<i>H. annuus</i>	X62352.1	42	CAA44224.1	43
Oleosin	<i>B. napus</i>	X82020.1	44	CAA57545.1	45
Oleosin	<i>Z. mays</i>	NM_001153560.1	46	NP_001147032.1	47
Oleosin	<i>O.sativa</i>	AAL40177.1	48	AAL40177.1	49
Oleosin	<i>B.oleracea</i>	AF117126.1	50	AAD24547.1	51
Oleosin	<i>C. arabica</i>	AY928084.1	52	AAY14574.1	53
Steroleosin	<i>S. indicum</i>	AAL13315	54	AAL13315	55
Steroleosin	<i>A. napus</i>	EU678274	56	ACG69522	57
Steroleosin	<i>Z. mays</i>	NM_001159142.1	58	NP_001152614.1	59
Steroleosin	<i>B. napus</i>	EF143915.1	60	ABM30178.1	61
Caloleosin	<i>S. indicum</i>	AF109921	62	AAF13743	63
Caloleosin	<i>G. max</i>	AF004809	64	AAB71227	65
Caloleosin	<i>Z. mays</i>	NM_001158434.1	66	NP_001151906	67

Caloleosin	<i>B. napus</i>	AY966447.1	68	AAV40837	69
Caloleosin	<i>C. revoluta</i>	FJ455154.1	70	ACJ70083	71
Caloleosin	<i>C. sativus</i>	EU232173.1	72	ABY56103.1	73

Oleosin, steroleosin and caloleosins are well known to those skilled in the art. Further sequences from many different species can be readily identified by methods well-known to those skilled in the art. For example, further sequences can be easily identified by an NCBI Entrez Cross-Database Search (available at <http://www.ncbi.nlm.nih.gov/sites/gquery>) using any one of the terms oleosin, steroleosin and caloleosin.

Plant lipids biosynthesis

All plant cells produce fatty acids from acetyl-CoA by a common pathway localized in plastids. Although a portion of the newly synthesized acyl chains is then used for lipid biosynthesis within the plastid (the prokaryotic pathway), a major portion is exported into the cytosol for glycerolipid assembly at the endoplasmic reticulum (ER) or other sites (the eukaryotic pathway). In addition, some of the extraplastidial glycerolipids return to the plastid, which results in considerable intermixing between the plastid and ER lipid pools (Ohlrogge and Jaworski 1997).

The simplest description of the plastidial pathway of fatty acid biosynthesis consists of two enzyme systems: acetyl-CoA carboxylase (ACCase) and fatty acid synthase (FAS). ACCase catalyzes the formation of malonyl-CoA from acetyl-CoA, and FAS transfers the malonyl moiety to acyl carrier protein (ACP) and catalyzes the extension of the growing acyl chain with malonyl-ACP.

The initial fatty acid synthesis reaction is catalyzed by 3-ketoacyl-ACP III (KAS III) which results in the condensation of acetyl-CoA and malonyl-ACP. Subsequent condensations are catalyzed by KAS I and KAS II. Before a subsequent cycle of fatty acid synthesis begins, the 3-ketoacyl-ACP intermediate is reduced to the saturated acyl-ACP in the remaining FAS reactions, catalyzed sequentially by the 3-ketoacyl-ACP reductase, 3 hydroxyacyl-ACP dehydrase, and the enoyl-ACP reductase.

The final products of FAS are usually 16:0 and 18:0-ACP, and the final fatty acid composition of a plant cell is in large part determined by activities of several enzymes that use these acyl-ACPs at the termination phase of fatty acid synthesis. Stearoyl-ACP desaturase modifies the final

product of FAS by insertion of a cis double bond at the 9 position of the C18:0-ACP. Reactions of fatty acid synthesis are terminated by hydrolysis or transfer of the acyl chain from the ACP. Hydrolysis is catalyzed by acyl-ACP thioesterases, of which there are two main types: one thioesterase relatively specific for 18:1-ACP and a second more specific for saturated acyl-ACPs.

5 Fatty acids that have been released from ACPs by thioesterases leave the plastid and enter into the eukaryotic lipid pathway, where they are primarily esterified to glycerolipids on the ER. Acyl transferases in the plastid, in contrast to thioesterases, terminate fatty acid synthesis by transesterifying acyl moieties from ACP to glycerol, and they are an essential part of the prokaryotic lipid pathway leading to plastid glycerolipid assembly.

10 *Triacylglycerol biosynthesis*

The only committed step in TAG biosynthesis is the last one, i.e. the addition of a third fatty acid to an existing diacylglycerol, thus generating TAG. In plants this step is predominantly (but not exclusively) performed by one of five (predominantly ER localised) TAG synthesising enzymes including: acyl CoA: diacylglycerol acyltransferase (DGAT1); an unrelated acyl CoA:

15 diacylglycerol acyl transferase (DGAT2); a soluble DGAT (DGAT3) which has less than 10% identity with DGAT1 or DGAT2 (Saha et al., 2006); phosphatidylcholine-sterol O-acyltransferase (PDAT); and a wax synthase (WSD1, Li et al., 2008). The DGAT1 and DGAT2 proteins are encoded by two distinct gene families, with DGAT1 containing approximately 500 amino acids and 10 predicted transmembrane domains and DGAT2 has only 320 amino acids

20 and two transmembrane domains (Shockey et al., 2006).

The term “triacylglycerol synthesising enzyme” or “TAG synthesising enzyme” as used herein means an enzyme capable of catalysing the addition of a third fatty acid to an existing diacylglycerol, thus generating TAG. Preferred TAG synthesising enzymes include but are not limited to: acyl CoA: diacylglycerol acyltransferase1 (DGAT1); diacylglycerol acyl transferase2

25 (DGAT2); phosphatidylcholine-sterol O-acyltransferase (PDAT) and cytosolic soluble form of DGAT (soluble DGAT or DGAT3).

Given that endogenous DGAT1 and DGAT2 appear to play roles in mature and senescing leaves (Kaup et al. 2002; Shockey et al. 2006), it is likely that plants possess a number of feedback mechanisms to control their activity. Indeed, Zou et al. (2008) recently identified a consensus

30 sequence (X-Leu-X-Lys-X-X-Ser-X-X-X-Val) within *Tropaeolum majus* (garden nasturtium) DGAT1 (TmDGAT1) sequences as a targeting motif typical of members of the SNF1-related protein kinase-1 (SnRK1) with Ser being the residue for phosphorylation. The SnRK1 proteins

are a class of Ser/Thr protein kinases that have been increasingly implicated in the global regulation of carbon metabolism in plants, e.g. the inactivation of sucrose phosphate synthase by phosphorylation (Halford & Hardie 1998). Zou *et al.* (2008) went on to demonstrate that the obliteration of a potential SnRK1 phosphorylation site in DGAT1 by single point mutation (Ser197Ala of TmDGAT1) led to the accumulation of significantly higher levels of TAG in the seed. This mutation increased activity by 38-80%, which led to a 20-50% increase in oil content on a per seed basis in Arabidopsis.

Phospholipid:DGA acyltransferase (PDAT) forms TAG from a molecule of phospholipid and a molecule of diacylglycerol. PDAT is quite active when expressed in yeast but does not appreciably increase TAG yields when expressed in plant seeds. PDAT and a proposed DAG:DAG transacylase are neutral lipid synthesizing enzymes that produce TAG, but are not considered part of the Kennedy Pathway.

A combination of wax ester synthase and DGAT enzyme (WS/DGAT) has been found in all neutral lipid producing prokaryotes studied so far. WS/DAGAT has extraordinary broad activity on a variety of unusual fatty acids, alcohols and even thiols. This enzyme has a putative membrane-spanning region but shows no sequence homology to the DGAT1 and DGAT2 families from eukaryotes or the WE synthase from jojoba (Jojoba is the only eukaryote that has been found to accumulate wax ester).

It should be noted that Lecithin-Cholesterol AcylTransferase (LCAT) and Acyl-coenzyme:Cholesterol AcylTransferase (ACAT) are enzymes that produce sterol esters (a form of neutral lipid) not TAGs.

In applications requiring the increase of neutral lipids evidence suggests that the higher activity and broader specificity of DGAT1 relative to DGAT2 is preferential. Where a specific fatty acid is preferred, such as a long-chain-PUFA, DGAT1 is still applicable, provided it accepts the fatty acid of choice. Plants generally incorporate long chain PUFAs in the sn-2 position. It is not known whether this is due to high activity of LPAT or low activity of DGAT1 on this substrate. For the improved specificity for PUFAs, a DGAT2 that prefers these fatty acids may be preferable, or the properties of DGAT1 could be altered using directed evolution or an equivalent procedure.

Examples of these TAG synthesising enzymes, suitable for use in the methods and compositions of the invention, from members of several plant species are provided in Table 2 below. The sequences (both polynucleotide and polypeptide are provided in the Sequence Listing)

Table 2

TAG synthesising enzyme	Species	cDNA accession no.	SEQ ID NO:	Protein accession no.	SEQ ID NO:
DGAT1	<i>A. thaliana</i>	NM_127503	74	NP_179535	75
DGAT1	<i>T. majus</i>	AY084052	76	AAM03340	77
DGAT1	<i>Z. mays</i>	EU039830	78	ABV91586	79
DGAT2	<i>A. thaliana</i>	NM_115011	80	NP_566952	81
DGAT2	<i>B. napus</i>	FJ858270	82	AC090187	83
DGAT3 (soluble DGAT)	<i>A. hypogaea</i>	AY875644	84	AAX62735	85
PDAT	<i>A. thaliana</i>	NM_121367	86	NP_196868	87
PDAT	<i>R. communis</i>	XM_00252130 4	88	XP_002521350	89

The inventions also contemplates use of modified TAG synthesizing enzymes, that are modified (for example in their sequence by substitutions, insertions or additions an the like) to alter their specificity and or activity.

5 TAG accumulation in leaves

A recent field survey of 302 angiosperm species in the north-central USA found that 24% have conspicuous cytosolic oil droplets in leaves, with usually one large oil droplet per mesophyll cell (Lersten *et al.*, 2006 [from Slocombe *et al* 2009]). The role of cytosolic leaf TAG is thought to be involved in carbon storage and/or membrane lipid re-modelling (for review see Slocombe *et al.*, 2009). Indeed, in senescing leaves, plastidial fatty acids are partitioned into TAG prior for further mobilization, and DGAT1 is thought to be instrumental in this process (Kaup *et al.*, 2002).

There have been several attempts to engineer plants to accumulate elevated levels of TAG in their leaves. The success of these has been somewhat limited by the relatively low level of TAG that accumulated and in some cases the majority of TAG accumulated in senescing leaves only, thus limiting the flexibility of harvesting and proportion of crop accumulating TAG at any one time (Bouvier-Nave *et al.*, 2001; Xu *et al.*, 2005; Winichayakul *et al.*, 2008; Andrianov *et al.*, 2010; Slocombe *et al.*, 2009 and references therein).

To date the attempts to accumulate TAG in leaves have predominantly focussed on three particular gene candidates including over expression of DGAT (TAG biosynthesis), mutation of TGD1 or CTS (resulting in the prevention of lipid remobilisation), and over expression of LEC1, LEC2 and WRI1 (transcriptional factors involved in storage oil and protein accumulation in developing seeds). Over expression of TAG and other neutral lipid synthesizing enzymes relies on the presence of sufficient substrate, in the expanding and or mature leaf this is assumed to be provided by the plastid (chloroplast in the case of the leaf) which synthesises lipids for membranes. In photosynthetic leaves of Arabidopsis it has been estimated that the turnover of membrane lipids is 4% of total fatty acids per day (Bao *et al.*, 2000). In senescing leaves, the existing plastidal membranes provide the bulk of fatty acids for partitioning into TAG prior to further mobilization.

Over-expression of the Arabidopsis DGAT1 gene in tobacco leaves results in enhanced TAG accumulation (Bouvier-Nave *et al.*, 2001), this was later repeated and quantified by Andrianov *et al.*, (2010). They calculated the TAG level increased 20 fold and lead to a doubling of lipid content from ~3% to ~6% of dry matter in mature leaves. A further increase to 6.8% was achieved by the over expression of LEC2 (a master regulator of seed maturation and seed oil storage) in mature leaves using the inducible Alc promoter (Andrianov *et al.*, 2010). No estimation of the extractable TAG was given, nor was there any calculation on the accumulation of TAG in expanding leaves.

Mutations in a permease-like protein TRIGALACTOSYLDIACYLGLYCEROL (TGD1), in Arabidopsis thaliana caused the accumulation of TAGs, oligogalactolipids and phosphatide; this was accompanied by a high incidence of embryo abortion and comparatively poor overall plant growth (Xu *et al.*, 2005).

Winichayakul *et al.*, (2008) over expressed Arabidopsis thaliana DGAT1 in the leaves of ryegrass (*Lolium perenne*) and found this lead to a 50% elevation of total extractable leaf lipid (from ~4% to 6% of dry matter). Furthermore, the elevated lipid level was present in new leaves

generated by repeated harvests spaced 2-3 weeks apart, indicating that the new emerging leaves were also capable of accumulating additional lipid. However, the elevated lipid level in these leaves typically began to decline to wild type levels when the leaves were more than 2 weeks old indicating that the lipids were being re-mobilised via catabolism (release from the glycerol backbone by lipase followed by β -oxidation).

Slocombe *et al.*, (2009) demonstrated that mutations in the CTS peroxisomal ABC transporter (cts-2) led to accumulation of up to 1.4% TAG in leaves, particularly during the onset of senescence. They also ectopically expressed LEC2 during senescence in the cts-2 background; while this did not elevate the overall accumulation of TAG over the cts-2 mutant it did increase the accumulation of seed oil type species of TAG in senescing tissue. While cts-2 blocks fatty acid breakdown it also led to a severe phenotype. Slocombe *et al.*, (2009) concluded that recycled membrane fatty acids may be able to be re-directed to TAG by expressing the seed-programme in senescing tissue or by a block in fatty acid breakdown.

Scott *et al.*, (2007) claimed that the co-expression of a triacylglyceride synthesising enzyme and polyoleosin (two or more oleosin units fused in a tandem head-to-tail arrangement) would enable the storage of lipid in a plant cell. Similarly, Cookson *et al.*, (2009) claimed that producing a single oleosin and a TAG synthesising enzyme within vegetative portions of a plant would lead to increased number of oil bodies and TAG in the vegetative tissue. Using either of these techniques leads to a maximum increase in lipid content (not necessarily in the form of TAG) of up to approximately 50%. Furthermore this level begins to decline as the leaves mature; typically in leaves greater than 2 weeks old (unpublished data).

Hence, the degree to which TAG can be accumulated in vegetative tissues appears to be limited to some extent by the fact that the endogenous fixed-carbon recovery machinery catabolises the TAG.

Leaf senescence – recycling of lipids via TAG intermediates

Leaf senescence is a highly controlled sequence of events leading ultimately to the death of cells, tissues and finally the whole organ. This entails regulated recruitment of nutrients together with their translocation from the senescing tissue to other tissues that are still growing and developing. The chloroplast is the first organelle of mesophyll cells to show symptoms of senescence and although breakdown of thylakoid membranes is initiated early in the leaf senescence cascade, the chloroplast envelope remains relatively intact until the very late stages of senescence. DGAT1 is up-regulated during senescence of Arabidopsis leaves and this is temporally correlated with

increased levels of TAG-containing fatty acids commonly found in chloroplast galactolipids. Recruitment of membrane carbon from senescing leaves, particularly senescing chloroplasts, to growing parts of the plant is a key feature of leaf senescence, and it involves de-esterification of thylakoid lipids and conversion of the resultant free fatty acids to phloem-mobile sucrose. De-
5 esterification of thylakoid lipids appears to be mediated by one or more senescence induced galactolipases. The formation of TAG appears to be an intermediate step in the mobilisation of membrane lipid carbon to phloem mobile sucrose during senescence (Kaup *et al.*, 2002).

Modified oleosins engineered to include artificially introduced cysteines

The modified oleosins of the invention, or for use in the methods of the invention, are modified
10 to contain at least one artificially introduced cysteine residue. Preferably the engineered oleosins contain at least two cysteines.

The encapsulation of the neutral lipids by oleosins containing engineered cysteines provides an alternative mechanism to accumulate appreciable quantities of TAG in leaves without the requirement to wait until senescence and without producing extreme phenotypes.

15 Various methods well-known to those skilled in the art may be used in production of the modified oleosins with artificially introduced cysteines.

Such methods include site directed mutagenesis (US 6,448,048) in which the polynucleotide encoding an oleosin is modified to introduce a cysteine into the encoded oleosin protein.

Alternatively the polynucleotide encoding the modified oleosins, may be synthesised in its entirety.

20 Further methodology for producing modified oleosins of the invention and for use in the methods of the invention, is provided in the Examples section.

The introduced cysteine may be an additional amino acid (i.e. an insertion) or may replace an existing amino acid (i.e. a replacement). Preferably the introduced cysteine replaces an existing amino acid. In a preferred embodiment the replaced amino acid is a charged residue. Preferably
25 the charged residue is predicted to be in the hydrophilic domains and therefore likely to be located on the surface of the oil body.

The hydrophilic, and hydrophobic regions/arms of the oleosin can be easily identified by those skilled in the art using standard methodology (for example: Kyte and Doolittle (1982)).

The modified oleosins of the invention are preferably range in molecular weight from 5 to 50 kDa, more preferably, 10 to 40kDa, more preferably 15 to 25 kDa.

The modified oleosins of the invention are preferably in the size range 100 to 300 amino acids, more preferably 110 to 260 amino acids, more preferably 120 to 250 amino acids, more
5 preferably 130 to 240 amino acids, more preferably 140 to 230 amino acids.

Preferably the modified oleosins comprise an N-terminal hydrophilic region, two centre hydrophobic regions (joined by a proline knot or knob) and a C-terminal hydrophilic region.

Preferably the modified oleosins can be divided almost equally their length into four parts which correspond to the N-terminal hydrophilic region (or arm), the two centre hydrophobic regions
10 (joined by a proline knot or knob) and a C-terminal hydrophilic region (or arm).

Preferably the topology of modified oleosin is attributed to its physical properties which include a folded hydrophobic core flanked by hydrophilic domains.

Preferably the modified oleosins can be formed into oil bodies when combined with triacylglycerol (TAG) and phospholipid.

15 Preferably topology confers an amphipathic nature to modified oleosin resulting in the hydrophobic domain being embedded in the phospholipid monolayer of the oil body while the flanking hydrophilic domains are exposed to the aqueous environment outside the oil body, such as in the cytoplasm.

In one embodiment the modified oleosin of the invention or used in the method of the invention,
20 comprises a sequence with at least 70% identity the hydrophobic domain of any of the oleosin protein sequences referred to in Table 1 above.

In one embodiment the modified oleosin of the invention or used in the method of the invention, comprises a sequence with at least 70% identity to any of the protein sequences referred to in Table 1 above.

25 In further embodiment the modified oleosin is essentially the same as any of the oleosins referred to in Table 1 above, apart from the additional artificially introduced cysteine or cysteines.

In a further embodiment the modified oleosin of the invention or used in the method of the invention, comprises a sequence with at least 70% identity to the oleosin sequence of SEQ ID NO: 16.

In further embodiment the modified oleosin has the same amino acid sequence as that of SEQ ID NO: 16, apart from the additional artificially introduced cysteine or cysteines.

In further embodiment the modified oleosin is has the amino acid sequence of any one of SEQ ID NO: 16 to 20.

5 *Fusion proteins with modified oleosins*

The invention also provides a fusion proteins including a modified oleosin of the invention fused to a protein of interest.

Preferably the protein of interest is at the N- or C-terminal end of the fusion protein.

10 Methods for recombinantly expressing fusion proteins are well known to those skilled in the art (Papapostolou and Howorka, 2009). Production of the fusion protein of the invention may typically involve fusing the coding sequence of the protein of interest to the coding sequence of the modified oleosin.

Such fusion proteins may be included in, or expressed in, the oil bodies of the invention and used to purify and deliver the protein of interest for a variety of applications, as discussed in Roberts
15 *et al*, (2008).

However in the invention makes it possible to take advantage of the option to vary the stability/integrity of the oil body provided by presence of the modified oleosins in the oil body, hence allowing for more stringent purification and delivery procedures.

Fusion proteins with un-modified oleosins

20 The invention also involves use of fusion protein including un-modified oleosin fused to a protein of interest. Production of the fusion protein of the invention may typically involve fusing the coding sequence of the protein of interest to the coding sequence of the un-modified oleosin.

Preferably the protein of interest is at the N- or C-terminal end of the fusion protein.

Such fusion proteins may be included or expressed in the oil bodies of the invention and used to
25 purify and deliver the protein of interest for a variety of applications, as discussed in Roberts *et al*, (2008).

The present invention however, takes advantage of the option to vary the stability/integrity of the oil body provided by presence of the modified oleosins in the oil body of the invention, hence allowing for more stringent purification and delivery procedures.

Vegetative tissues

- 5 Vegetative tissue include, shoots, leaves, roots, stems. A preferred vegetative tissue is a leaf.

Vegetative tissue specific promoters

An example of a vegetative specific promoter is found in US 6,229,067; and US 7,629,454; and US 7,153,953; and US 6,228,643.

Pollen specific promoters

- 10 An example of a pollen specific promoter is found in US 7,141,424; and US 5,545,546; and US 5,412,085; and US 5,086,169; and US 7,667,097.

Seed specific promoters

An example of a seed specific promoter is found in US 6,342,657; and US 7,081,565; and US 7,405,345; and US 7,642,346; and US 7,371,928.

- 15 *Fruit specific promoters*

An example of a fruit specific promoter is found in US 5,536,653; and US 6,127,179; and US 5,608,150; and US 4,943,674.

Polynucleotides and fragments

- 20 The term "polynucleotide(s)," as used herein, means a single or double-stranded deoxyribonucleotide or ribonucleotide polymer of any length but preferably at least 15 nucleotides, and include as non-limiting examples, coding and non-coding sequences of a gene, sense and antisense sequences complements, exons, introns, genomic DNA, cDNA, pre-mRNA, mRNA, rRNA, siRNA, miRNA, tRNA, ribozymes, recombinant polypeptides, isolated and purified naturally occurring DNA or RNA sequences, synthetic RNA and DNA sequences,
25 nucleic acid probes, primers and fragments.

A "fragment" of a polynucleotide sequence provided herein is a subsequence of contiguous nucleotides that is capable of specific hybridization to a target of interest, e.g., a sequence that is

[illegible]

more preferably at least 93 nucleotides, more preferably at least 94 nucleotides, more preferably at least 95 nucleotides, more preferably at least 96 nucleotides, more preferably at least 97 nucleotides, more preferably at least 98 nucleotides, more preferably at least 99 nucleotides, more preferably at least 100 nucleotides, more preferably at least 150 nucleotides, more preferably at least 200 nucleotides, more preferably at least 250 nucleotides, more preferably at least 300 nucleotides, more preferably at least 350 nucleotides, more preferably at least 400 nucleotides, more preferably at least 450 nucleotides and most preferably at least 500 nucleotides of contiguous nucleotides of a polynucleotide disclosed. A fragment of a polynucleotide sequence can be used in antisense, RNA interference (RNAi), gene silencing, triple helix or ribozyme technology, or as a primer, a probe, included in a microarray, or used in polynucleotide-based selection methods of the invention.

The term "primer" refers to a short polynucleotide, usually having a free 3'OH group, that is hybridized to a template and used for priming polymerization of a polynucleotide complementary to the target.

The term "probe" refers to a short polynucleotide that is used to detect a polynucleotide sequence that is complementary to the probe, in a hybridization-based assay. The probe may consist of a "fragment" of a polynucleotide as defined herein.

Polypeptides and fragments

The term "polypeptide", as used herein, encompasses amino acid chains of any length but preferably at least 5 amino acids, including full-length proteins, in which amino acid residues are linked by covalent peptide bonds. Polypeptides of the present invention, or used in the methods of the invention, may be purified natural products, or may be produced partially or wholly using recombinant or synthetic techniques. The term may refer to a polypeptide, an aggregate of a polypeptide such as a dimer or other multimer, a fusion polypeptide, a polypeptide fragment, a polypeptide variant, or derivative thereof.

A "fragment" of a polypeptide is a subsequence of the polypeptide that performs a function that is required for the biological activity and/or provides three dimensional structure of the polypeptide. The term may refer to a polypeptide, an aggregate of a polypeptide such as a dimer or other multimer, a fusion polypeptide, a polypeptide fragment, a polypeptide variant, or derivative thereof capable of performing the above enzymatic activity.

The term "isolated" as applied to the polynucleotide or polypeptide sequences disclosed herein is used to refer to sequences that are removed from their natural cellular environment. An isolated molecule may be obtained by any method or combination of methods including biochemical, recombinant, and synthetic techniques.

- 5 The term "recombinant" refers to a polynucleotide sequence that is removed from sequences that surround it in its natural context and/or is recombined with sequences that are not present in its natural context.

A "recombinant" polypeptide sequence is produced by translation from a "recombinant" polynucleotide sequence.

- 10 The term "derived from" with respect to polynucleotides or polypeptides of the invention being derived from a particular genera or species, means that the polynucleotide or polypeptide has the same sequence as a polynucleotide or polypeptide found naturally in that genera or species. The polynucleotide or polypeptide, derived from a particular genera or species, may therefore be produced synthetically or recombinantly.

15 *Variants*

- As used herein, the term "variant" refers to polynucleotide or polypeptide sequences different from the specifically identified sequences, wherein one or more nucleotides or amino acid residues is deleted, substituted, or added. Variants may be naturally occurring allelic variants, or non-naturally occurring variants. Variants may be from the same or from other species and may
- 20 encompass homologues, paralogues and orthologues. In certain embodiments, variants of the inventive polypeptides and polypeptides possess biological activities that are the same or similar to those of the inventive polypeptides or polypeptides. The term "variant" with reference to polypeptides and polypeptides encompasses all forms of polypeptides and polypeptides as defined herein.

25 *Polynucleotide variants*

- Variant polynucleotide sequences preferably exhibit at least 50%, more preferably at least 51%, more preferably at least 52%, more preferably at least 53%, more preferably at least 54%, more preferably at least 55%, more preferably at least 56%, more preferably at least 57%, more preferably at least 58%, more preferably at least 59%, more preferably at least 60%, more
- 30 preferably at least 61%, more preferably at least 62%, more preferably at least 63%, more

preferably at least 64%, more preferably at least 65%, more preferably at least 66%, more preferably at least 67%, more preferably at least 68%, more preferably at least 69%, more preferably at least 70%, more preferably at least 71%, more preferably at least 72%, more preferably at least 73%, more preferably at least 74%, more preferably at least 75%, more preferably at least 76%, more preferably at least 77%, more preferably at least 78%, more preferably at least 79%, more preferably at least 80%, more preferably at least 81%, more preferably at least 82%, more preferably at least 83%, more preferably at least 84%, more preferably at least 85%, more preferably at least 86%, more preferably at least 87%, more preferably at least 88%, more preferably at least 89%, more preferably at least 90%, more preferably at least 91%, more preferably at least 92%, more preferably at least 93%, more preferably at least 94%, more preferably at least 95%, more preferably at least 96%, more preferably at least 97%, more preferably at least 98%, and most preferably at least 99% identity to a sequence of the present invention. Identity is found over a comparison window of at least 20 nucleotide positions, preferably at least 50 nucleotide positions, more preferably at least 100 nucleotide positions, and most preferably over the entire length of a polynucleotide of the invention.

Polynucleotide sequence identity can be determined in the following manner. The subject polynucleotide sequence is compared to a candidate polynucleotide sequence using BLASTN (from the BLAST suite of programs, version 2.2.5 [Nov 2002]) in bl2seq (Tatiana A. Tatusova, Thomas L. Madden (1999), "Blast 2 sequences - a new tool for comparing protein and nucleotide sequences", FEMS Microbiol Lett. 174:247-250), which is publicly available from NCBI (<ftp://ftp.ncbi.nih.gov/blast/>). The default parameters of bl2seq are utilized except that filtering of low complexity parts should be turned off.

The identity of polynucleotide sequences may be examined using the following unix command line parameters:

```
bl2seq -i nucleotideseq1 -j nucleotideseq2 -F F -p blastn
```

The parameter -F F turns off filtering of low complexity sections. The parameter -p selects the appropriate algorithm for the pair of sequences. The bl2seq program reports sequence identity as both the number and percentage of identical nucleotides in a line "Identities = ".

Polynucleotide sequence identity may also be calculated over the entire length of the overlap between a candidate and subject polynucleotide sequences using global sequence alignment programs (e.g. Needleman, S. B. and Wunsch, C. D. (1970) J. Mol. Biol. 48, 443-453). A full

implementation of the Needleman-Wunsch global alignment algorithm is found in the needle program in the EMBOSS package (Rice, P. Longden, I. and Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite, Trends in Genetics June 2000, vol 16, No 6. pp.276-277) which can be obtained from <http://www.hgmp.mrc.ac.uk/Software/EMBOSS/>. The European Bioinformatics Institute server also provides the facility to perform EMBOSS-needle global alignments between two sequences on line at <http://www.ebi.ac.uk/emboss/align/>.

Alternatively the GAP program may be used which computes an optimal global alignment of two sequences without penalizing terminal gaps. GAP is described in the following paper: Huang, X. (1994) On Global Sequence Alignment. Computer Applications in the Biosciences 10, 227-235.

- 10 A preferred method for calculating polynucleotide % sequence identity is based on aligning sequences to be compared using Clustal X (Jeanmougin et al., 1998, Trends Biochem. Sci. 23, 403-5.)

Polynucleotide variants of the present invention also encompass those which exhibit a similarity to one or more of the specifically identified sequences that is likely to preserve the functional equivalence of those sequences and which could not reasonably be expected to have occurred by random chance. Such sequence similarity with respect to polypeptides may be determined using the publicly available bl2seq program from the BLAST suite of programs (version 2.2.5 [Nov 2002]) from NCBI (<ftp://ftp.ncbi.nih.gov/blast/>).

The similarity of polynucleotide sequences may be examined using the following unix command line parameters:

```
bl2seq -i nucleotideseq1 -j nucleotideseq2 -F F -p tblastx
```

The parameter -F F turns off filtering of low complexity sections. The parameter -p selects the appropriate algorithm for the pair of sequences. This program finds regions of similarity between the sequences and for each such region reports an "E value" which is the expected number of times one could expect to see such a match by chance in a database of a fixed reference size containing random sequences. The size of this database is set by default in the bl2seq program. For small E values, much less than one, the E value is approximately the probability of such a random match.

Variant polynucleotide sequences preferably exhibit an E value of less than 1×10^{-6} more preferably less than 1×10^{-9} , more preferably less than 1×10^{-12} , more preferably less than $1 \times$

10⁻¹⁵, more preferably less than 1 x 10⁻¹⁸, more preferably less than 1 x 10⁻²¹, more preferably less than 1 x 10⁻³⁰, more preferably less than 1 x 10⁻⁴⁰, more preferably less than 1 x 10⁻⁵⁰, more preferably less than 1 x 10⁻⁶⁰, more preferably less than 1 x 10⁻⁷⁰, more preferably less than 1 x 10⁻⁸⁰, more preferably less than 1 x 10⁻⁹⁰ and most preferably less than 1 x 10⁻¹⁰⁰ when compared with any one of the specifically identified sequences.

Alternatively, variant polynucleotides of the present invention, or used in the methods of the invention, hybridize to the specified polynucleotide sequences, or complements thereof under stringent conditions.

The term "hybridize under stringent conditions", and grammatical equivalents thereof, refers to the ability of a polynucleotide molecule to hybridize to a target polynucleotide molecule (such as a target polynucleotide molecule immobilized on a DNA or RNA blot, such as a Southern blot or Northern blot) under defined conditions of temperature and salt concentration. The ability to hybridize under stringent hybridization conditions can be determined by initially hybridizing under less stringent conditions then increasing the stringency to the desired stringency.

With respect to polynucleotide molecules greater than about 100 bases in length, typical stringent hybridization conditions are no more than 25 to 30° C (for example, 10° C) below the melting temperature (T_m) of the native duplex (see generally, Sambrook *et al.*, Eds, 1987, Molecular Cloning, A Laboratory Manual, 2nd Ed. Cold Spring Harbor Press; Ausubel *et al.*, 1987, Current Protocols in Molecular Biology, Greene Publishing.). T_m for polynucleotide molecules greater than about 100 bases can be calculated by the formula $T_m = 81.5 + 0.41\% (G + C - \log(Na^+))$. (Sambrook *et al.*, Eds, 1987, Molecular Cloning, A Laboratory Manual, 2nd Ed. Cold Spring Harbor Press; Bolton and McCarthy, 1962, PNAS 84:1390). Typical stringent conditions for polynucleotide of greater than 100 bases in length would be hybridization conditions such as prewashing in a solution of 6X SSC, 0.2% SDS; hybridizing at 65°C, 6X SSC, 0.2% SDS overnight; followed by two washes of 30 minutes each in 1X SSC, 0.1% SDS at 65° C and two washes of 30 minutes each in 0.2X SSC, 0.1% SDS at 65°C.

With respect to polynucleotide molecules having a length less than 100 bases, exemplary stringent hybridization conditions are 5 to 10° C below T_m. On average, the T_m of a polynucleotide molecule of length less than 100 bp is reduced by approximately (500/oligonucleotide length)⁰ C.

With respect to the DNA mimics known as peptide nucleic acids (PNAs) (Nielsen *et al.*, Science. 1991 Dec 6;254(5037):1497-500) T_m values are higher than those for DNA-DNA or DNA-RNA

hybrids, and can be calculated using the formula described in Giesen *et al.*, Nucleic Acids Res. 1998 Nov 1;26(21):5004-6. Exemplary stringent hybridization conditions for a DNA-PNA hybrid having a length less than 100 bases are 5 to 10° C below the T_m.

Variant polynucleotides of the present invention, or used in the methods of the invention, also encompasses polynucleotides that differ from the sequences of the invention but that, as a consequence of the degeneracy of the genetic code, encode a polypeptide having similar activity to a polypeptide encoded by a polynucleotide of the present invention. A sequence alteration that does not change the amino acid sequence of the polypeptide is a "silent variation". Except for ATG (methionine) and TGG (tryptophan), other codons for the same amino acid may be changed by art recognized techniques, e.g., to optimize codon expression in a particular host organism.

Polynucleotide sequence alterations resulting in conservative substitutions of one or several amino acids in the encoded polypeptide sequence without significantly altering its biological activity are also included in the invention. A skilled artisan will be aware of methods for making phenotypically silent amino acid substitutions (see, e.g., Bowie *et al.*, 1990, Science 247, 1306).

Variant polynucleotides due to silent variations and conservative substitutions in the encoded polypeptide sequence may be determined using the publicly available bl2seq program from the BLAST suite of programs (version 2.2.5 [Nov 2002]) from NCBI (<ftp://ftp.ncbi.nih.gov/blast/>) via the tblastx algorithm as previously described.

Polypeptide variants

The term "variant" with reference to polypeptides encompasses naturally occurring, recombinantly and synthetically produced polypeptides. Variant polypeptide sequences preferably exhibit at least 50%, more preferably at least 51%, more preferably at least 52%, more preferably at least 53%, more preferably at least 54%, more preferably at least 55%, more preferably at least 56%, more preferably at least 57%, more preferably at least 58%, more preferably at least 59%, more preferably at least 60%, more preferably at least 61%, more preferably at least 62%, more preferably at least 63%, more preferably at least 64%, more preferably at least 65%, more preferably at least 66%, more preferably at least 67%, more preferably at least 68%, more preferably at least 69%, more preferably at least 70%, more preferably at least 71%, more preferably at least 72%, more preferably at least 73%, more preferably at least 74%, more preferably at least 75%, more preferably at least 76%, more preferably at least 77%, more preferably at least 78%, more preferably at least 79%, more preferably at least 80%, more preferably at least 81%, more preferably at least 82%, more

preferably at least 83%, more preferably at least 84%, more preferably at least 85%, more preferably at least 86%, more preferably at least 87%, more preferably at least 88%, more preferably at least 89%, more preferably at least 90%, more preferably at least 91%, more preferably at least 92%, more preferably at least 93%, more preferably at least 94%, more preferably at least 95%, more preferably at least 96%, more preferably at least 97%, more preferably at least 98%, and most preferably at least 99% identity to a sequences of the present invention. Identity is found over a comparison window of at least 20 amino acid positions, preferably at least 50 amino acid positions, more preferably at least 100 amino acid positions, and most preferably over the entire length of a polypeptide of the invention.

10 Polypeptide sequence identity can be determined in the following manner. The subject polypeptide sequence is compared to a candidate polypeptide sequence using BLASTP (from the BLAST suite of programs, version 2.2.5 [Nov 2002]) in bl2seq, which is publicly available from NCBI (<ftp://ftp.ncbi.nih.gov/blast/>). The default parameters of bl2seq are utilized except that filtering of low complexity regions should be turned off.

15 Polypeptide sequence identity may also be calculated over the entire length of the overlap between a candidate and subject polynucleotide sequences using global sequence alignment programs. EMBOSS-needle (available at <http://www.ebi.ac.uk/emboss/align/>) and GAP (Huang, X. (1994) On Global Sequence Alignment: Computer Applications in the Biosciences 10, 227-235.) as discussed above are also suitable global sequence alignment programs for calculating
20 polypeptide sequence identity.

A preferred method for calculating polypeptide % sequence identity is based on aligning sequences to be compared using Clustal X (Jeanmougin et al., 1998, Trends Biochem. Sci. 23, 403-5.)

Polypeptide variants of the present invention, or used in the methods of the invention, also
25 encompass those which exhibit a similarity to one or more of the specifically identified sequences that is likely to preserve the functional equivalence of those sequences and which could not reasonably be expected to have occurred by random chance. Such sequence similarity with respect to polypeptides may be determined using the publicly available bl2seq program from the BLAST suite of programs (version 2.2.5 [Nov 2002]) from NCBI
30 (<ftp://ftp.ncbi.nih.gov/blast/>). The similarity of polypeptide sequences may be examined using the following unix command line parameters:

```
bl2seq -i peptideseq1 -j peptideseq2 -F F -p blastp
```

Variant polypeptide sequences preferably exhibit an E value of less than 1×10^{-6} more preferably less than 1×10^{-9} , more preferably less than 1×10^{-12} , more preferably less than 1×10^{-15} , more preferably less than 1×10^{-18} , more preferably less than 1×10^{-21} , more preferably less than 1×10^{-30} , more preferably less than 1×10^{-40} , more preferably less than 1×10^{-50} , more preferably less than 1×10^{-60} , more preferably less than 1×10^{-70} , more preferably less than 1×10^{-80} , more preferably less than 1×10^{-90} and most preferably 1×10^{-100} when compared with any one of the specifically identified sequences.

The parameter -F F turns off filtering of low complexity sections. The parameter -p selects the appropriate algorithm for the pair of sequences. This program finds regions of similarity between the sequences and for each such region reports an "E value" which is the expected number of times one could expect to see such a match by chance in a database of a fixed reference size containing random sequences. For small E values, much less than one, this is approximately the probability of such a random match.

Conservative substitutions of one or several amino acids of a described polypeptide sequence without significantly altering its biological activity are also included in the invention. A skilled artisan will be aware of methods for making phenotypically silent amino acid substitutions (see, e.g., Bowie et al., 1990, Science 247, 1306).

Constructs, vectors and components thereof

The term "genetic construct" refers to a polynucleotide molecule, usually double-stranded DNA, which may have inserted into it another polynucleotide molecule (the insert polynucleotide molecule) such as, but not limited to, a cDNA molecule. A genetic construct may contain the necessary elements that permit transcribing the insert polynucleotide molecule, and, optionally, translating the transcript into a polypeptide. The insert polynucleotide molecule may be derived from the host cell, or may be derived from a different cell or organism and/or may be a recombinant polynucleotide. Once inside the host cell the genetic construct may become integrated in the host chromosomal DNA. The genetic construct may be linked to a vector.

The term "vector" refers to a polynucleotide molecule, usually double stranded DNA, which is used to transport the genetic construct into a host cell. The vector may be capable of replication in at least one additional host system, such as *E. coli*.

The term "expression construct" refers to a genetic construct that includes the necessary elements that permit transcribing the insert polynucleotide molecule, and, optionally, translating the transcript into a polypeptide. An expression construct typically comprises in a 5' to 3' direction:

- a) a promoter functional in the host cell into which the construct will be transformed,
- 5 b) the polynucleotide to be expressed, and
- c) a terminator functional in the host cell into which the construct will be transformed.

The term "coding region" or "open reading frame" (ORF) refers to the sense strand of a genomic DNA sequence or a cDNA sequence that is capable of producing a transcription product and/or a polypeptide under the control of appropriate regulatory sequences. The coding sequence may, in some cases, identified by the presence of a 5' translation start codon and a 3' translation stop codon. When inserted into a genetic construct, a "coding sequence" is capable of being expressed when it is operably linked to promoter and terminator sequences.

"Operably-linked" means that the sequenced to be expressed is placed under the control of regulatory elements that include promoters, tissue-specific regulatory elements, temporal regulatory elements, enhancers, repressors and terminators.

The term "noncoding region" refers to untranslated sequences that are upstream of the translational start site and downstream of the translational stop site. These sequences are also referred to respectively as the 5' UTR and the 3' UTR. These regions include elements required for transcription initiation and termination, mRNA stability, and for regulation of translation efficiency.

Terminators are sequences, which terminate transcription, and are found in the 3' untranslated ends of genes downstream of the translated sequence. Terminators are important determinants of mRNA stability and in some cases have been found to have spatial regulatory functions.

The term "promoter" refers to nontranscribed cis-regulatory elements upstream of the coding region that regulate gene transcription. Promoters comprise cis-initiator elements which specify the transcription initiation site and conserved boxes such as the TATA box, and motifs that are bound by transcription factors. Introns within coding sequences can also regulate transcription and influence post-transcriptional processing (including splicing, capping and polyadenylation).

A promoter may be homologous with respect to the polynucleotide to be expressed. This means that the promoter and polynucleotide are found operably linked in nature.

Alternatively the promoter may be heterologous with respect to the polynucleotide to be expressed. This means that the promoter and the polynucleotide are not found operably linked in nature.

A "transgene" is a polynucleotide that is taken from one organism and introduced into a different organism by transformation. The transgene may be derived from the same species or from a different species as the species of the organism into which the transgene is introduced.

An "inverted repeat" is a sequence that is repeated, where the second half of the repeat is in the complementary strand, e.g.,

(5')GATCTA.....TAGATC(3')

(3')CTAGAT.....ATCTAG(5')

Read-through transcription will produce a transcript that undergoes complementary base-pairing to form a hairpin structure provided that there is a 3-5 bp spacer between the repeated regions.

Host cells

Host cells may be derived from, for example, bacterial, fungal, yeast, insect, mammalian, algal or plant organisms. Host cells may also be synthetic cells. Preferred host cells are eukaryotic cells. A particularly preferred host cell is a plant cell, particularly a plant cell in a vegetative tissue of a plant.

A "transgenic plant" refers to a plant which contains new genetic material as a result of genetic manipulation or transformation. The new genetic material may be derived from a plant of the same species as the resulting transgenic plant or from a different species.

Methods for isolating or producing polynucleotides

The polynucleotide molecules of the invention can be isolated by using a variety of techniques known to those of ordinary skill in the art. By way of example, such polypeptides can be isolated through use of the polymerase chain reaction (PCR) described in Mullis et al., Eds. 1994 The Polymerase Chain Reaction, Birkhauser, incorporated herein by reference. The polypeptides of the invention can be amplified using primers, as defined herein, derived from the polynucleotide sequences of the invention.

Further methods for isolating polynucleotides of the invention include use of all, or portions of, the polypeptides having the sequence set forth herein as hybridization probes. The technique of hybridizing labelled polynucleotide probes to polynucleotides immobilized on solid supports such as nitrocellulose filters or nylon membranes, can be used to screen the genomic or cDNA libraries. Exemplary hybridization and wash conditions are: hybridization for 20 hours at 65°C in 5.0 X SSC, 0.5% sodium dodecyl sulfate, 1 X Denhardt's solution; washing (three washes of twenty minutes each at 55°C) in 1.0 X SSC, 1% (w/v) sodium dodecyl sulfate, and optionally one wash (for twenty minutes) in 0.5 X SSC, 1% (w/v) sodium dodecyl sulfate, at 60°C. An optional further wash (for twenty minutes) can be conducted under conditions of 0.1 X SSC, 1% (w/v) sodium dodecyl sulfate, at 60°C.

The polynucleotide fragments of the invention may be produced by techniques well-known in the art such as restriction endonuclease digestion, oligonucleotide synthesis and PCR amplification.

A partial polynucleotide sequence may be used, in methods well-known in the art to identify the corresponding full length polynucleotide sequence. Such methods include PCR-based methods, 5'RACE (Frohman MA, 1993, *Methods Enzymol.* 218: 340-56) and hybridization-based method, computer/database-based methods. Further, by way of example, inverse PCR permits acquisition of unknown sequences, flanking the polynucleotide sequences disclosed herein, starting with primers based on a known region (Triglia et al., 1998, *Nucleic Acids Res* 16, 8186, incorporated herein by reference). The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template. Divergent primers are designed from the known region. In order to physically assemble full-length clones, standard molecular biology approaches can be utilized (Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Ed. Cold Spring Harbor Press, 1987).

It may be beneficial, when producing a transgenic plant from a particular species, to transform such a plant with a sequence or sequences derived from that species. The benefit may be to alleviate public concerns regarding cross-species transformation in generating transgenic organisms. Additionally when down-regulation of a gene is the desired result, it may be necessary to utilise a sequence identical (or at least highly similar) to that in the plant, for which reduced expression is desired. For these reasons among others, it is desirable to be able to identify and isolate orthologues of a particular gene in several different plant species.

Variants (including orthologues) may be identified by the methods described.

*Methods for identifying variants**Physical methods*

Variant polypeptides may be identified using PCR-based methods (Mullis et al., Eds. 1994 The Polymerase Chain Reaction, Birkhauser). Typically, the polynucleotide sequence of a primer,
5 useful to amplify variants of polynucleotide molecules of the invention by PCR, may be based on a sequence encoding a conserved region of the corresponding amino acid sequence.

Alternatively library screening methods, well known to those skilled in the art, may be employed (Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed. Cold Spring Harbor Press, 1987). When identifying variants of the probe sequence, hybridization and/or wash stringency
10 will typically be reduced relatively to when exact sequence matches are sought.

Polypeptide variants may also be identified by physical methods, for example by screening expression libraries using antibodies raised against polypeptides of the invention (Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed. Cold Spring Harbor Press, 1987) or by identifying polypeptides from natural sources with the aid of such antibodies.

Computer based methods
15

The variant sequences of the invention, including both polynucleotide and polypeptide variants, may also be identified by computer-based methods well-known to those skilled in the art, using public domain sequence alignment algorithms and sequence similarity search tools to search sequence databases (public domain databases include Genbank, EMBL, Swiss-Prot, PIR and
20 others). See, e.g., Nucleic Acids Res. 29: 1-10 and 11-16, 2001 for examples of online resources. Similarity searches retrieve and align target sequences for comparison with a sequence to be analyzed (i.e., a query sequence). Sequence comparison algorithms use scoring matrices to assign an overall score to each of the alignments.

An exemplary family of programs useful for identifying variants in sequence databases is the
25 BLAST suite of programs (version 2.2.5 [Nov 2002]) including BLASTN, BLASTP, BLASTX, tBLASTN and tBLASTX, which are publicly available from (<ftp://ftp.ncbi.nih.gov/blast/>) or from the National Center for Biotechnology Information (NCBI), National Library of Medicine, Building 38A, Room 8N805, Bethesda, MD 20894 USA. The NCBI server also provides the facility to use the programs to screen a number of publicly available sequence databases.
30 BLASTN compares a nucleotide query sequence against a nucleotide sequence database.

BLASTP compares an amino acid query sequence against a protein sequence database. BLASTX compares a nucleotide query sequence translated in all reading frames against a protein sequence database. tBLASTN compares a protein query sequence against a nucleotide sequence database dynamically translated in all reading frames. tBLASTX compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database. The BLAST programs may be used with default parameters or the parameters may be altered as required to refine the screen.

The use of the BLAST family of algorithms, including BLASTN, BLASTP, and BLASTX, is described in the publication of Altschul et al., *Nucleic Acids Res.* 25: 3389-3402, 1997.

- 10 The "hits" to one or more database sequences by a queried sequence produced by BLASTN, BLASTP, BLASTX, tBLASTN, tBLASTX, or a similar algorithm, align and identify similar portions of sequences. The hits are arranged in order of the degree of similarity and the length of sequence overlap. Hits to a database sequence generally represent an overlap over only a fraction of the sequence length of the queried sequence.
- 15 The BLASTN, BLASTP, BLASTX, tBLASTN and tBLASTX algorithms also produce "Expect" values for alignments. The Expect value (E) indicates the number of hits one can "expect" to see by chance when searching a database of the same size containing random contiguous sequences. The Expect value is used as a significance threshold for determining whether the hit to a database indicates true similarity. For example, an E value of 0.1 assigned to a polynucleotide hit is
20 interpreted as meaning that in a database of the size of the database screened, one might expect to see 0.1 matches over the aligned portion of the sequence with a similar score simply by chance. For sequences having an E value of 0.01 or less over aligned and matched portions, the probability of finding a match by chance in that database is 1% or less using the BLASTN, BLASTP, BLASTX, tBLASTN or tBLASTX algorithm.
- 25 Multiple sequence alignments of a group of related sequences can be carried out with CLUSTALW (Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22:4673-4680, <http://www-igbmc.u-strasbg.fr/BioInfo/ClustalW/Top.html>) or T-COFFEE (Cedric
30 Notredame, Desmond G. Higgins, Jaap Heringa, T-Coffee: A novel method for fast and accurate multiple sequence alignment, *J. Mol. Biol.* (2000) 302: 205-217)) or PILEUP, which uses progressive, pairwise alignments. (Feng and Doolittle, 1987, *J. Mol. Evol.* 25, 351).

Pattern recognition software applications are available for finding motifs or signature sequences. For example, MEME (Multiple Em for Motif Elicitation) finds motifs and signature sequences in a set of sequences, and MAST (Motif Alignment and Search Tool) uses these motifs to identify similar or the same motifs in query sequences. The MAST results are provided as a series of
5 alignments with appropriate statistical data and a visual overview of the motifs found. MEME and MAST were developed at the University of California, San Diego.

PROSITE (Bairoch and Bucher, 1994, Nucleic Acids Res. 22, 3583; Hofmann *et al.*, 1999, Nucleic Acids Res. 27, 215) is a method of identifying the functions of uncharacterized proteins translated from genomic or cDNA sequences. The PROSITE database (www.expasy.org/prosite)
10 contains biologically significant patterns and profiles and is designed so that it can be used with appropriate computational tools to assign a new sequence to a known family of proteins or to determine which known domain(s) are present in the sequence (Falquet *et al.*, 2002, Nucleic Acids Res. 30, 235). Prosearch is a tool that can search SWISS-PROT and EMBL databases with a given sequence pattern or signature.

15 *Methods for isolating polypeptides*

The polypeptides of the invention, or used in the methods of the invention, including variant polypeptides, may be prepared using peptide synthesis methods well known in the art such as direct peptide synthesis using solid phase techniques (e.g. Stewart *et al.*, 1969, in Solid-Phase Peptide Synthesis, WH Freeman Co, San Francisco California, or automated synthesis, for
20 example using an Applied Biosystems 431A Peptide Synthesizer (Foster City, California). Mutated forms of the polypeptides may also be produced during such syntheses.

The polypeptides and variant polypeptides of the invention, or used in the methods of the invention, may also be purified from natural sources using a variety of techniques that are well known in the art (e.g. Deutscher, 1990, Ed, Methods in Enzymology, Vol. 182, Guide to Protein
25 Purification,).

Alternatively the polypeptides and variant polypeptides of the invention, or used in the methods of the invention, may be expressed recombinantly in suitable host cells and separated from the cells as discussed below.

Methods for producing constructs and vectors

The genetic constructs of the present invention comprise one or more polynucleotide sequences of the invention and/or polynucleotides encoding polypeptides of the invention, and may be useful for transforming, for example, bacterial, fungal, insect, mammalian or plant organisms.

- 5 The genetic constructs of the invention are intended to include expression constructs as herein defined.

Methods for producing and using genetic constructs and vectors are well known in the art and are described generally in Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2nd Ed. Cold Spring Harbor Press, 1987 ; Ausubel *et al.*, Current Protocols in Molecular Biology, Greene
10 Publishing, 1987).

Methods for producing host cells comprising polynucleotides, constructs or vectors

The invention provides a host cell which comprises a genetic construct or vector of the invention.

Host cells comprising genetic constructs, such as expression constructs, of the invention are useful in methods well known in the art (e.g. Sambrook *et al.*, Molecular Cloning : A Laboratory
15 Manual, 2nd Ed. Cold Spring Harbor Press, 1987 ; Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing, 1987) for recombinant production of polypeptides of the invention. Such methods may involve the culture of host cells in an appropriate medium in conditions suitable for or conducive to expression of a polypeptide of the invention. The
20 expressed recombinant polypeptide, which may optionally be secreted into the culture, may then be separated from the medium, host cells or culture medium by methods well known in the art (e.g. Deutscher, Ed, 1990, Methods in Enzymology, Vol 182, Guide to Protein Purification).

Methods for producing plant cells and plants comprising constructs and vectors

The invention further provides plant cells which comprise a genetic construct of the invention, and plant cells modified to alter expression of a polynucleotide or polypeptide of the invention,
25 or used in the methods of the invention. Plants comprising such cells also form an aspect of the invention.

Methods for transforming plant cells, plants and portions thereof with polypeptides are described in Draper *et al.*, 1988, Plant Genetic Transformation and Gene Expression. A Laboratory
Manual, Blackwell Sci. Pub. Oxford, p. 365; Potrykus and Spangenburg, 1995, Gene Transfer to
30 Plants. Springer-Verlag, Berlin.; and Gelvin *et al.*, 1993, Plant Molecular Biol. Manual. Kluwer

Acad. Pub. Dordrecht. A review of transgenic plants, including transformation techniques, is provided in Galun and Breiman, 1997, Transgenic Plants. Imperial College Press, London.

Methods for genetic manipulation of plants

A number of plant transformation strategies are available (e.g. Birch, 1997, Ann Rev Plant Phys Plant Mol Biol, 48, 297, Hellens RP, et al (2000) Plant Mol Biol 42: 819-32, Hellens R et al Plant Meth 1: 13). For example, strategies may be designed to increase expression of a polynucleotide/polypeptide in a plant cell, organ and/or at a particular developmental stage where/when it is normally expressed or to ectopically express a polynucleotide/polypeptide in a cell, tissue, organ and/or at a particular developmental stage which/when it is not normally expressed. The expressed polynucleotide/polypeptide may be derived from the plant species to be transformed or may be derived from a different plant species.

Transformation strategies may be designed to reduce expression of a polynucleotide/polypeptide in a plant cell, tissue, organ or at a particular developmental stage which/when it is normally expressed. Such strategies are known as gene silencing strategies.

Genetic constructs for expression of genes in transgenic plants typically include promoters for driving the expression of one or more cloned polynucleotide, terminators and selectable marker sequences to detect presence of the genetic construct in the transformed plant.

The promoters suitable for use in the constructs of this invention are functional in a cell, tissue or organ of a monocot or dicot plant and include cell-, tissue- and organ-specific promoters, cell cycle specific promoters, temporal promoters, inducible promoters, constitutive promoters that are active in most plant tissues, and recombinant promoters. Choice of promoter will depend upon the temporal and spatial expression of the cloned polynucleotide, so desired. The promoters may be those normally associated with a transgene of interest, or promoters which are derived from genes of other plants, viruses, and plant pathogenic bacteria and fungi. Those skilled in the art will, without undue experimentation, be able to select promoters that are suitable for use in modifying and modulating plant traits using genetic constructs comprising the polynucleotide sequences of the invention. Examples of constitutive plant promoters include the CaMV 35S promoter, the nopaline synthase promoter and the octopine synthase promoter, and the Ubi 1 promoter from maize. Plant promoters which are active in specific tissues, respond to internal developmental signals or external abiotic or biotic stresses are described in the scientific literature. Exemplary promoters are described, e.g., in WO 02/00894, which is herein incorporated by reference.

Exemplary terminators that are commonly used in plant transformation genetic construct include, e.g., the cauliflower mosaic virus (CaMV) 35S terminator, the *Agrobacterium tumefaciens* nopaline synthase or octopine synthase terminators, the *Zea mays* zein gene terminator, the *Oryza sativa* ADP-glucose pyrophosphorylase terminator and the *Solanum tuberosum* PI-II terminator.

- 5 Selectable markers commonly used in plant transformation include the neomycin phosphotransferase II gene (NPT II) which confers kanamycin resistance, the aadA gene, which confers spectinomycin and streptomycin resistance, the phosphinothricin acetyl transferase (*bar* gene) for Ignite (AgrEvo) and Basta (Hoechst) resistance, and the hygromycin phosphotransferase gene (*hpt*) for hygromycin resistance.
- 10 Use of genetic constructs comprising reporter genes (coding sequences which express an activity that is foreign to the host, usually an enzymatic activity and/or a visible signal (e.g., luciferase, GUS, GFP) which may be used for promoter expression analysis in plants and plant tissues are also contemplated. The reporter gene literature is reviewed in Herrera-Estrella *et al.*, 1993, Nature 303, 209, and Schrott, 1995, In: Gene Transfer to Plants (Potrykus, T., Spangenberg, Eds)
- 15 Springer Verlag, Berline, pp. 325-336.

The following are representative publications disclosing genetic transformation protocols that can be used to genetically transform the following plant species: Rice (Alam *et al.*, 1999, Plant Cell Rep. 18, 572); apple (Yao *et al.*, 1995, Plant Cell Reports 14, 407-412); maize (US Patent Serial Nos. 5, 177, 010 and 5, 981, 840); wheat (Ortiz *et al.*, 1996, Plant Cell Rep. 15, 1996, 20 877); tomato (US Patent Serial No. 5, 159, 135); potato (Kumar *et al.*, 1996 Plant J. 9, : 821); cassava (Li *et al.*, 1996 Nat. Biotechnology 14, 736); lettuce (Michelmore *et al.*, 1987, Plant Cell Rep. 6, 439); tobacco (Horsch *et al.*, 1985, Science 227, 1229); cotton (US Patent Serial Nos. 5, 846, 797 and 5, 004, 863); grasses (US Patent Nos. 5, 187, 073 and 6, 020, 539); peppermint (Niu *et al.*, 1998, Plant Cell Rep. 17, 165); citrus plants (Pena *et al.*, 1995, Plant Sci.104, 183); 25 caraway (Krens *et al.*, 1997, Plant Cell Rep. 17, 39); banana (US Patent Serial No. 5, 792, 935); soybean (US Patent Nos. 5, 416, 011 ; 5, 569, 834 ; 5, 824, 877 ; 5, 563, 04455 and 5, 968, 830); pineapple (US Patent Serial No. 5, 952, 543); poplar (US Patent No. 4, 795, 855); monocots in general (US Patent Nos. 5, 591, 616 and 6, 037, 522); brassica (US Patent Nos. 5, 188, 958 ; 5, 463, 174 and 5, 750, 871); cereals (US Patent No. 6, 074, 877); pear (Matsuda *et al.*, 2005, Plant 30 Cell Rep. 24(1):45-51); Prunus (Ramesh *et al.*, 2006 Plant Cell Rep. 25(8):821-8; Song and Sink 2005 Plant Cell Rep. 2006 ;25(2):117-23; Gonzalez Padilla *et al.*, 2003 Plant Cell Rep.22(1):38-45); strawberry (Oosumi *et al.*, 2006 Planta. 223(6):1219-30; Folta *et al.*, 2006 Planta Apr 14; PMID: 16614818), rose (Li *et al.*, 2003), Rubus (Graham *et al.*, 1995 Methods Mol Biol.

1995;44:129-33), tomato (Dan et al., 2006, Plant Cell Reports V25:432-441), apple (Yao et al., 1995, *Plant Cell Rep.* **14**, 407-412), Canola (*Brassica napus* L.)(Cardoza and Stewart, 2006 Methods Mol Biol. 343:257-66), safflower (Orlikowska et al, 1995, Plant Cell Tissue and Organ Culture 40:85-91), ryegrass (Altpeter et al, 2004 Developments in Plant Breeding 11(7):255-250), rice (Christou et al, 1991 Nature Biotech. 9:957-962), maize (Wang et al 2009 In: Handbook of Maize pp. 609-639) and *Actinidia eriantha* (Wang et al.; 2006, Plant Cell Rep. 25,5: 425-31). Transformation of other species is also contemplated by the invention. Suitable methods and protocols are available in the scientific literature.

Plants

- 10 The term “plant” is intended to include a whole plant, any part of a plant, a seed, a fruit, propagules and progeny of a plant.

The term ‘propagule’ means any part of a plant that may be used in reproduction or propagation, either sexual or asexual, including seeds and cuttings.

- 15 The plants of the invention may be grown and either self-ed or crossed with a different plant strain and the resulting hybrids, with the desired phenotypic characteristics, may be identified. Two or more generations may be grown to ensure that the subject phenotypic characteristics are stably maintained and inherited. Plants resulting from such standard breeding approaches also form an aspect of the present invention.

Abbreviations

- 20 **Oleosin (or Ole)_0-0 means an oleosin without engineered cysteines.**
Oleosin (or Ole)_1-1 means an oleosin with one engineered cysteine in each hydrophilic arm.
Oleosin (or Ole)_1-3 means an oleosin with one engineered cysteine in the N-terminal hydrophilic arm and three engineered cysteines in the C-terminal hydrophilic arm.
25 **Oleosin (or Ole)_3-1 means an oleosin with three engineered cysteines in the N-terminal hydrophilic arm and one engineered cysteine in the C-terminal hydrophilic arm.**
Oleosin (or Ole)_3-3 means an oleosin with three engineered cysteines in the N-terminal hydrophilic arm and three engineered cysteines in the C-terminal hydrophilic arm.
Oleosin (or Ole)_5-6 means an oleosin with five engineered cysteines in the N-terminal
30 **hydrophilic arm and six engineered cysteines in the C-terminal hydrophilic arm.**
Oleosin (or Ole)_6-7 means an oleosin with six engineered cysteines in the N-terminal hydrophilic arm and seven engineered cysteines in the C-terminal hydrophilic arm.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows the sequence of the Oleosin_0-0 and DGAT1 (S205A) construct. CaMV35 is the Cauliflower Mosaic Virus 35S promoter. attB1 is the GATEWAY™ recombination site. UBQ10 is the intron from the A. thaliana UBQ10 gene. OCS terminator is the octopine synthase terminator.

Figure 2 shows the Oleosin_1-1 and DGAT1 (S205A) construct arrangement, as transformed into Arabidopsis thaliana.

Figure 3 shows the sequence of the Oleosin_1-3 and DGAT1 (S205A) construct. CaMV35 is the Cauliflower Mosaic Virus 35S promoter. attB1 is the GATEWAY™ recombination site. UBQ10 is the intron from the A. thaliana UBQ10 gene. OCS terminator is the octopine synthase terminator.

Figure 4 shows the Oleosin_3-1 and DGAT1 (S205A) construct. CaMV35 is the Cauliflower Mosaic Virus 35S promoter. attB1 is the GATEWAY™ recombination site. UBQ10 is the intron from the A. thaliana UBQ10 gene. OCS terminator is the octopine synthase terminator.

Figure 5 shows the Oleosin_3-3 and DGAT1 (S205A) construct. CaMV35 is the Cauliflower Mosaic Virus 35S promoter. attB1 is the GATEWAY™ recombination site. UBQ10 is the intron from the A. thaliana UBQ10 gene. OCS terminator is the octopine synthase terminator.

Figure 6 shows a map of the construct pRSh1 used for transforming plants. The map shows the arrangement of the oleosins, with artificially introduced cysteines (in this case Oleo_3-3) under the control of the CaMV35s promoter as well as Arabidopsis thaliana DGAT1 (S205A) also under the control of the CaMV35s promoter. Other oleosin sequences and TAG synthesising enzyme sequences can of course be substituted for Oleo_3-3 and DGAT1 respectively.

Figure 7 shows dot blot comparison of anti-sesame seed oleosin antibodies binding to purified recombinant sesame seed oleosin with and without engineered cysteine residues.

Figure 8 shows immunoblot analysis to detect *E. coli* expressed oleosin cysteine proteins in AOBs. Equal volume of AOB (7.5 µL including 2x SDS loading dye without reducing agent) was loaded per lane. The mM concentration of GSSG is indicated above each lane.

Figure 9 shows SDS and SDS-UREA PAGE/immunoblot analysis of *E. coli* expressed Ole-0-0, Ole-1-1 and Ole-3-3. Samples were prepared from inclusion bodies (IB) and artificial oil bodies

(AOBs) in the presence and absence of reducing agents (DTT and β -ME) or oxidising agent (GSSG), where equal amounts of protein were loaded in adjacent lanes.

Figure 10 shows immunoblot analysis of oleosin (Oleo_0-0, Oleo_1-3, Oleo_3-1, and Oleo_3-3, SEQ ID NOs 11-20) accumulation in the seeds of transgenic *Arabidopsis thaliana* expressing both DGAT1 (S205A) and a sesame oleosin under the control of CaMV35S promoters.

Figure 11 shows immunoblot analysis of oleosin (Oleo_0-0, Oleo_1-3, Oleo_3-1, and Oleo_3-3, SEQ ID NOs 11-20) accumulation in the oil bodies of transgenic *Arabidopsis thaliana* expressing both DGAT1 (S205A) and a sesame oleosin under the control of CaMV35S promoters. The appearance of the oligomeric oleosin bands (dimeric and trimeric) in the presence of oxidising agent (+) indicates the disulfide bonds are able to form on the outside of native oil bodies.

Figure 12 shows immunoblot analysis of oleosin (Oleo_0-0, Oleo_1-3, Oleo_3-1, and Oleo_3-3, SEQ ID NOs 11-20) accumulation in the leaves of transgenic *Arabidopsis thaliana* expressing both DGAT1 (S205A) and a sesame oleosin under the control of CaMV35S promoters.

Figure 13 shows immunoblot of recombinant oleosin accumulation (black arrow) in transgenic *Arabidopsis* leaves.

Figure 14 shows FAMES GC/MS results demonstrating accumulation of additional lipids (black arrows) in *Arabidopsis* leaves over expressing DGAT1 (S205A) and Ole_3,3.

Figure 15 shows GC/MS results for total leaf lipid profile of wild type and independent lines of transgenic *Arabidopsis* containing DGAT1 (S205A) and Ole_3,3. Grey arrow indicates internal standard. Black arrows indicate additional neutral lipids (wax esters, sterol esters and TAGs. Open arrows show three lines (41S, 18A and 47C) which accumulate substantial quantities of neutral lipids in their leaves compared to wild type (and line 50A).

Figure 16 shows GC/MS results showing total TAG profile of wild type and transgenic *Arabidopsis* (containing DGAT1 (S205A) and Ole_3,3) 2, 3, 4 and 5 weeks after germination. Black arrows indicate additional TAGs found in transgenic leaves but not wild type.

Figure 17 shows FAMES GC/MS results showing total leaf lipid profiles of wild type and transgenic *Trifolium repens* (containing DGAT1 (S205A) and Ole_3,3).

Figure 18 shows FAMES GC/MS results showing C18:1 and C18:2 leaf lipid profiles of wild type and transgeneic *Trifolium repens* (containing DGAT1 (S205A) and Ole_3,3).

EXAMPLES

This invention will now be illustrated with reference to the following non-limiting examples.

5 EXAMPLE 1: Creating rabbit anti-sesame seed oleosin antibodies

Generating rabbit anti-sesame seed oleosin antibodies

Full length sesame seed oleosin containing a C-terminal His tag (nucleotide sequence is shown in SEQ ID NO: 1) was expressed in *E. coli* and inclusion bodies were prepared by standard techniques. The inclusion bodies were solubilised in Binding Buffer (100 mM phosphate buffer
10 pH 8.0, 500mM NaCl, 8M urea and 10 mM imidazole) and loaded onto a column containing equilibrated ion metal affinity chromatography (IMAC) Ni agarose (Invitrogen). Non-bound proteins were removed from the column by washing with 6 volumes of Wash Buffer (100 mM phosphate buffer pH 8.0, 500mM NaCl, 6M urea and 50 mM imidazole). Protein was eluted in 1 vol. aliquots of Elution Buffer (100 mM phosphate buffer pH 8.0, 500 mM NaCl,
15 6M urea and 250mM imidazole). Eluted fractions were analysed by SDS-PAGE/Coomassie stain and the protein concentration measured using the Bradford's Assay. 265µg of the IMAC-purified recombinant oleosin protein was mixed with an equal amount of Freund's Complete Adjuvant to a final volume of 0.5mL. Following collection of the pre-bleed, the first injection was administered into multiple sites across the back of the neck and shoulder area of a
20 rabbit. Booster shots containing 77µg of the purified oleosin were delivered at three and seven weeks after the primary, and a test bleed of ~3mL was removed for preliminary analysis at nine weeks. Serum was preserved by the addition of 0.25% v/v phenol and 0.01% v/v merthiolate, and stored in 200µL aliquots at -20°C.

The sensitivity of the rabbit anti-sesame seed oleosin antibodies was evaluated by immuno-dotting which indicated that 0.25ng of sesame seed oleosin could be regularly detected with a
25 1/2,000 dilution of the antibody (Figure 7).

EXAMPLE 2: Design and *E. coli* expression of modified oleosins containing one or more artificially introduced cysteine residue

Construct design for expression in E. coli

A number of modified oleosin constructs for expression in *E. coli* were designed. These
5 contained either one or three cysteine residues on the N-terminal and C-terminal hydrophilic arms. The constructs were based on the nucleotide sequence and translated polypeptide sequence from a sesame seed oleosin, GenBank clone AF091840 which contains no cysteine residues (SEQ ID NO: 16).

All clones were subcloned into pET29b using engineered NdeI/XhoI sites. In addition, a ProTrp
10 coding sequence was added to the coding region of the 3' end of the C-terminal hydrophilic arm to mimic the amino acid residues encoded for by the NcoI site previously engineered by Peng et al (2006) Stability enhancement of native and artificial oil bodies by genipin crosslink. Taiwan Patent I 250466.

Oleosin-cysteine proteins mutated to include cysteine residues in both the N- and C- terminal
15 hydrophilic regions described here are designated Ole-1-1, Ole-1-3, Ole-3-1, and Ole-3-3 (SEQ ID NO 2, 3, 4, and 5 respectively), where the first and the second numeral digits correspond to the number of disulfide bonds in the N- and C- terminus, respectively. The standard oleosin without the cysteine residues was used as a control and was designated as Ole-0-0 (SEQ ID NO 1).

20 The cysteines were substituted for charged residues predicted to be on the surface of the oil bodies and are listed below.

N-terminal single cysteine (Ole-1-x) Glu3Cys

N-terminal triple cysteine (Ole-3-x) Glu3Cys Arg12Cys Gln23Cys

C-terminal single cysteine (Ole-x-1) Gln137Cys

25 C-terminal triple cysteine (Ole-x-3) Gln112Cys Lys123Cys Gln137Cys

The constructs were designed so could be relatively simply sub cloned from the GENEART provided backbone (pCR4Blunt-TOPO) into pET29b (Novogen) via NcoI/XhoI digestion and ligation. This placed the oleosin coding sequence downstream of the pET29 N-terminal S•tag

fusion and upstream of the C-terminal His tag (Figures 1-5 and SEQ ID Nos 1-10). The oleosin and modified oleosin sequences used are summarised in the Summary of Sequences table.

Expression in E. coli and purification of modified oleosins containing at least one artificially introduced cysteine

- 5 Expression of the recombinant sesame seed oleosins (with and without engineered cysteines) in the *E. coli* expression system was evaluated by SDS-PAGE/Coomassie brilliant blue staining and SDS-PAGE/immunoblot analysis using antibodies raised against the sesame seed oleosin (described in Example 1).

Expression of recombinant modified oleosin was induced in a freshly inoculated 10mL culture of
10 *E. coli* (BL21 Rosetta-Gami) containing an oleosin (with or without engineered cysteine residues) coding sequence in the pET29 expression vector. The culture was grown at 37°C, 220rpm, until mid log phase ($OD_{600} 0.5 - 0.7$); expression was induced by the addition of IPTG to 1 mM final concentration. The induced culture was incubated at 37°C, 220rpm, for a further 2-3 h. Given the properties of modified oleosin the applicants did not attempt to express it in a
15 soluble form but rather chose to extract it from inclusion bodies. Aliquots (1mL) of the culture were transferred to 1.5mL microfuge tubes and the cells pelleted by centrifugation ($2655 \times g$ for 5 min at 4°C).

Pelleted cells were resuspended in BugBuster® Reagent (Merck) at 5 mL/g of wet cell pellet, with the addition of DNase to 40 µg/mL and mixed gently on a rotator for 30 min followed by
20 centrifugation at 8000g for 10 min at 4°C. The resultant cell pellet was retreated with BugBuster® and DNase as above. The remaining soluble protein and suspended cell debris was separated from the insoluble inclusion bodies by centrifugation at 8000g for 10 min at 4°C.

Recombinant oleosins were further purified from the inclusion bodies using a procedure adapted from D'Andréa *et al.* (2007). Briefly: the inclusion body preparation was washed by re-
25 suspension in 200 mM sodium carbonate buffer pH 11 (5 mL per gram of original cell pellet) and re-pelleted by centrifugation at $8000 \times g$ for 10 min at 4°C. The washed inclusion body pellet was again re-suspended in 5 mL 200 mM sodium carbonate buffer per gram of pellet and added to 9 volumes of freshly prepared chloroform:methanol mix (5:4 v/v) giving a final ratio of 5:4:1 (chloroform:methanol:buffer). The suspension was gently mixed for 5 min which formed a
30 milky, single phase mixture; this was centrifuged at $10,000 \times g$ for 10 min at 4°C, and the supernatant containing the modified oleosin was carefully separated from the pellet and

transferred into a new tube. Aliquots of the supernatant were dried down under a stream of nitrogen and the protein re-solubilised in 8M urea and quantified by Qubit™ (Invitrogen).

EXAMPLE 3: Use of anti-sesame seed oleosin antibodies to bind sesame seed oleosin with artificially introduced cysteines

5 A dot-blot was used to compare the ability of the anti-sesame seed oleosin antibodies (Abs) described in Example 1 to bind to oleosin without cysteines versus the oleosins containing cysteines (described in Example 2). Dilution series from 12 to 0.25ng of purified Ole-0-0, Ole-1-3 and Ole-3-1 were spotted onto a pre-equilibrated Hybond-P PVDF Transfer membrane. This was incubated with the anti-sesame seed oleosin antibodies at 1:2000 as the primary Ab. The blot was then incubated with the appropriate secondary Ab and developed by chemiluminescence (Figure 7). The results indicate that on an immunoblot, the anti-sesame seed oleosin antibodies are up to an order of magnitude more sensitive to the oleosin without cysteine residues than the oleosins with cysteine residues. As a consequence of the different sensitivities it was necessary to load different quantities of recombinant protein onto the gels for analysis by immunoblotting. 10 Despite the non uniform lane loading it is still possible to compare different oleosins between lanes in terms of their relative distribution between monomeric and oligomeric forms.

EXAMPLE 4: Creation of artificial oil bodies with *E. coli* expressed modified oleosins containing at least one artificially introduced cysteine and altering the degree of cross linking

20 *Preparation of artificial oil bodies*

Artificial oil bodies (AOBs) were then prepared by drying down aliquots of the supernatant described in Example 3, calculated to contain either 150µg or 1mg of recombinant oleosin.

The process of generating AOBs involved combining PL, TAG, and the recombinant oleosin/modified oleosin. In the absence of strong chaotropic agents the disruptive force required to dissociate individual recombinant oleosins from the purified fraction involved several alternating cycles of sonicating and cooling. This was achieved by solubilising the 150µg and 1mg oleosin/modified oleosin samples in 20µL chloroform containing 150µg PL (Sigma, Cat#P3644) and mixed with 60µL of purified sesame seed oil (Tzen and Huang 1992) and 940µL of AOB buffer (50mM sodium phosphate buffer pH 8.0, 100mM NaCl). The complete mixture was then sonicated three times for 30sec (Sonics & Materials Vibra-Cell VC600, 30 600 W, 20 kHz; 1/8" tapered micro-tip probe, power setting #3).

The applicants also found that the purification procedure could be successfully scaled up and when a 50g cell pellet was used as the starting material it was necessary to substitute the stream of nitrogen with a rotary vacuum evaporator to remove the chloroform and the majority of the methanol. At this point the majority of oleosin/modified oleosin precipitated out of the azeotropic solvent and was separated by centrifugation at 12,000g for 10min.

Inclusion bodies were suspended in 1mL AOB Buffer II (50 mM sodium phosphate, pH 8.0, 100 mM NaCl, 20 mM β -mercaptoethanol, 10 mM DTT and 5% [v/v] sesame oil) and then sonicated 4x. AOBs were concentrated by centrifugation at 12,000 rpm for 10min, this resulted in the formation of a suspension of AOBs overlaying the aqueous fraction. The underlying aqueous fraction was removed by pipette, and the remaining AOBs were washed (to remove soluble proteins and reducing agents) by gentle agitation in 1mL AOB Buffer III (50 mM sodium phosphate, pH 8.0, 100 mM NaCl). After washing, the AOBs were re-concentrated by centrifugation, and the underlying aqueous fraction removed, then re-suspended by vortexing in AOB Buffer IV (50 mM sodium phosphate buffer, pH 8.0, 100 mM NaCl, 1 mM GSSG) and the AOBs stored at 4°C for further analyses.

Recombinant Ole-0-0, and all variations of the oleosin-cysteines were successfully expressed and located in *E. coli* inclusion bodies (Figure 9). Ole-0-0 was predominantly present as a monomer (in both inclusion bodies as well as AOBs); this migrated fractionally faster than the 20kDa molecular weight marker (in reducing and non reducing SDS and SDS-UREA PAGE). Also present were two slower migrating immunoreactive bands of approximately 35 and 36 kDa which likely correspond to two forms of dimeric oleosin. While Ole-0-0 is not predicted to contain any cysteine residues the overall intensity and ratio of the two apparent dimers was influenced by the presence of reducing agents (β -ME @ 5% of the sample loading buffer and 10mM DTT).

In the inclusion bodies, the predominant form of Ole-1-1 is monomeric. Only one dimeric form appeared to be present and this was not influenced by reducing agents or urea. Ole-1-1 from AOBs (generated in the presence of reducing agent and then in the presence of oxidising agent) showed a large increase in the ratio of dimer to monomer as well as the formation of trimeric, tetrameric and likely pentameric oligomers (the electrophoretic focus of these oligomers was considerably improved in the SDS-UREA gel). Removal of the GSSG and re-introduction of reducing agents to the AOBs resulted in the presence of only monomer and dimer in similar proportions seen in the inclusion bodies. AOBs generated with Ole-1-1 (in the absence of both

reducing agents and GSSG) showed the presence of almost equal portions of monomer and dimer and a small amount of trimer, indicating that the conditions under which the AOBs are formed have some reducing potential. The subsequent addition of GSSG resulted in an increase in the oligomeric portions as well as the appearance of a tetrameric form.

5 While the monomer was the predominant form of Ole-3-3 in the inclusion bodies, a comparatively high percentage was also present in multiple oligomeric forms. The proportion of oligomers declined to a small extent with the addition of reducing agent and slightly more by the addition of both reducing and chaotropic agents. Oligomeric forms of Ole-3-3 that were larger than a trimer were poorly resolved when the recombinant protein was extracted from AOBs. The
10 creation of large oligomeric forms was promoted by the addition of GSSG and in the absence of reducing and chaotropic agents a portion of these oligomeric forms failed to enter the stacking gel. Combined, these results indicate that on the AOBs, Ole-3-3 was highly cross-linked and the position of the cross-links was more variable compared to the Ole-3-3 recovered from the inclusion bodies. This suggests that, despite considerable pre-existing cross-linking (within the
15 inclusion bodies), on the AOB Ole-3-3 has access to a high number of potential partners for cross-linking. Similarly for Ole-1-3 and Ole-3-1, the number of cross-linked species increased when there was more than one cysteine on one or both hydrophilic regions (Figures 8 and 9).

It could be anticipated that in non-reducing SDS-PAGE, oligomers containing the same number of oleosins but with the disulfide bonds in different positions would migrate differently to each
20 other. Indeed this can be seen in Figure 8 where the data indicates that the position of the oleosin arms, relative to one another are at different positions over the oil body. For example the Ole-1-1 can only form one disulfide bond per arm and this has to form at the same position, where as the presence of three cysteines enables more than one disulfide bond to form but it also allows the disulfide bonds to form with different degrees of hydrophilic arm overlap as well as having
25 multiple oleosins bound to the same arm (Figures 8 and 9).

The addition of SDS and reducing agents (DTT and β -ME) decreased the number of oligomeric complexes (Figure 9). The addition of SDS and urea results in a similar pattern to SDS alone except that the previously resolved multiple dimeric forms migrated as one and the trimeric and tetrameric forms appear to be in higher abundance presumably because they are also migrating as
30 single bands which increases intensity correspondingly (Figure 9). In contrast, the presence of SDS, reducing agent and urea resulted in fewer oligomeric forms of Ole-1-1 and Ole-1-3 but not Ole-3-1 or Ole-3-3 (Figure 9). In the case of Ole-3-1 and Ole-3-3 it appears that the urea does not completely denature the disulfides oleosins and may indeed prevent the complete reduction

of the disulfide bonds. It could be that these bonds are formed during the generation of inclusion bodies (would need to see reduced and non reduced inclusion body preps). Furthermore, the presence of the dimeric oleosin formed in the absence of engineered cysteine residues (Figures 8 and 9) indicates that some oligomerisation is due to other types of attraction, e.g, strong hydrophobic bonding that is not fully disrupted by SDS but can be almost completely disrupted by the combination of both SDS and urea (Figure 8 and 9).

The effect of increasing the number of potential cross-linking sites in an oleosin peptide on AOB integrity and emulsion stability can be assessed as follows.

Quantitative Determination of AOB integrity

- 10 Assessment of AOB stability and integrity using either absorbance (OD₆₀₀), direct counting of AOBs using a hemocytometer, or visual evaluation of coalescence by microscopy proved to be highly variable and amongst other things was influenced by the: degree of pre-sampling agitation; quantity of sample removed; time left under the microscope. To avoid this the applicants devised a simple method to quantify the amount of TAG released from the AOBs into the surrounding media during a variety of treatments as a means of comparing integrity.
- 15 Essentially equal quantities (based on FAMES-GC/MS estimation of TAG and Bradfords determination of protein) of AOB preparations are made up to a total volume of 200μL using AOB buffer (containing Proteinase K [PNK] when appropriate at a 1:1 ratio of PNK:total proteins in OB or AOB samples in a 250μL GC glass insert tubes and covered with a plastic cap.
- 20 Following the treatment (elevated temperature or exposure to PNK) 15μL of fish oil (Vitamax®, Australia) is added to the sample and mixed by vortexing followed by centrifugation at 5,200g for 1min. The addition of fish oil followed by vortexing enables any TAG that had leaked from the AOBs to mix with the added fish oil and be floated by brief centrifugation. 4μL of the oil phase is sampled and subjected to fatty acid methyl esterification (FAME) and then analysed by
- 25 GC-MS (Shimadzu model numbers, fitted with a 50mQC2/BPX70-0.25 GC capillary column (SGE) as described by Browse *et al.* (1986). In the absence of added fish oil the quantity of TAG that had leaked from the AOBs was too small to form a samplable visible layer even after centrifugation; in such a case the maximum volume would have been 6μL. The very different lipid profiles of fish oil and sesame oil enabled us to easily distinguish the leaked TAG from the
- 30 added TAG.

Using the internal C15:0 and C17:0 standards the applicants can calculate the absolute amounts of C18:2 (the major lipid in sesame seed oil) recovered after treatment.

Determination of AOB integrity and emulsion stability at elevated temperature

Oil in water emulsions are less stable at elevated temperatures; hence, it is of interest to investigate if modified oleosins with varying numbers in introduced cysteines influence AOB integrity at elevated temperature. To achieve this the applicants determine the integrity (using the method described above) of OBs and AOBs (containing different oleosins) in a phosphate buffer (50mM Na-phosphate buffer pH8, 100mM NaCl) at 95°C. AOBs are heated for 2h. Integrity is determined as above.

The effect of higher ratios of crosslinked oleosin:TAG on the stability of AOBs in rumen fluid can be assessed as follows.

10 Determination of AOB integrity in rumen fluid

One of the aims of disulfide was to provide some degree of protection from biohydrogenation by rumen microflora. Assessment of AOB stability with rumen fluid can be assessed as follows. AOBs are added to an equal volume (25µL) of rumen fluid. Samples are incubated at 39°C for 0, 15, 30, 60, 120 and 240min, at the end of the incubation an equal volume of loading buffer (Invitrogen) is added, mixed and heated at 70°C for 10min. 15µL of each sample/loading buffer mix is compared by SDS-PAGE/immunoblot. Integrity is determined as above.

Analysis of AOB integrity in Proteinase K

To investigate the influence of modified oleosin in a controllable and repeatable highly degradative environment integrity is determined (using the method described above) of AOBs (containing different modified oleosins) after incubation in an phosphate buffer (50mM Na-phosphate buffer pH8, 100mM NaCl) containing 1:1 (g/g protein) Proteinase K (Invitrogen) at 37°C for 4h. While the maximum activity of Proteinase K is achieved below 65°C the lower temperature is used in order to reduce the influence of temperature on AOB instability. Integrity is determined as above.

25 EXAMPLE 5: Design and *in planta* expression of modied oleosin containing one or more artificially introduced cysteines*Construct design for expression in planta*

The applicants synthesised individual coding sequences for the sesame seed oleosin (based on GenBank clone AF091840) with different numbers of cysteines in the N- and C-terminal arms.

The coding sequence was flanked by a 5' NotI site and a 3' NdeI site. A separate acceptor cassette was synthesised containing an attL1 site, a NotI site and NdeI site followed by a nos termination sequence, a forward facing CaMV35s promoter, the *Arabidopsis thaliana* DGAT1 (S205A) (SEQ ID NOs 11-20 and Figures 1-5) plus its own UBQ10 intron, an attL2 site. The
5 sesame seed oleosins with different numbers of cysteines were individually transferred to the acceptor cassette via the NotI and NdeI sites. Each of these completed cassettes were then transferred to a plant binary vector pRSh1, Figure 6 (Winichayakul et al., 2008) via the LR recombination reaction. This placed the oleosin downstream of a CaMV35S promoter (already contained within pRSh1) and placed a nos terminator (already contained within pRSh1)
10 downstream of the Arabidopsis DGAT1 (S205A) (Figures 1-5). The nucleotide sequences encoding the sesame seed oleosins (with cysteines) and DGAT1 were optimised for expression in *Arabidopsis thaliana*, this included optimisation of codon frequency, GC content, removal of cryptic splice sites, removal of mRNA instability sequences, removal of potential polyadenylation recognition sites, and addition of tetranucleotide stop codon (Brown et al, 1990;
15 Beelman and Parker, 1995; Rose, 2004; Rose and Beliakoff, 2000; Norris et al., 1993).

It should be noted that the oleosin sequence used is for example only. Any oleosin or steroleosin or caoleosin sequences could be engineered to contain cross-linking regions. The coding sequences of the complete ORFs (after splicing) were checked against repeat of the original oleosin translated sequence and found to be identical over the oleosin coding regions.

20 Transformation of *Arabidopsis thaliana* with sesame seed oleosins containing cysteines

Transformation of *Arabidopsis thaliana* var Columbia (with constructs described above), analyses of T2 seeds for modified oleosin, immunoblot analysis of *Arabidopsis thaliana* oil bodies containing sesame seed oleosin with different numbers of cysteines was performed as described previously (Scott et al., 2007).

25 Both the floral-dip (Clough, 1998) and floral-drop methods (Martinez-Trujillo, 2004) were used in the transformation of *Arabidopsis* by *Agrobacterium tumefaciens* GV3101 containing the binary constructs. T1 seed was collected from the treated plants, germinated and selected by spraying at 14 d and 21 d post-germination with Basta®. Basta® resistant T1 plants (71, 62 and 23 transformants containing the single sesame seed oleosin, and modified oleosin constructs
30 respectively) were transplanted, allowed to self-fertilise, set seed and the T2 seed was collected. Equal quantities of seed extract from Basta® resistant *Arabidopsis* plants were analysed by SDS-PAGE/immunoblot with the anti-sesame seed oleosin antibodies; recombinant sesame seed

oleosin and modified oleosin of the appropriate size was observed in the majority of samples (Figure 10). Southern blot analysis was performed on selected T2 lines to determine the number of insertion sites.

EXAMPLE 6: Extraction and purification oil bodies with modified oleosins containing at

5 least one artificially introduced cysteine from the seeds of *Arabidopsis thaliana*

Crude Oil Body Preparations from Arabidopsis thaliana seeds

Crude OB preparations were prepared, from seed of plants produced as described in Example 5, by either grinding 200mg seed with a mortar and pestle containing a spatula tip of sand and 750µL Extraction Buffer (10mM phosphate buffer, pH 7.5 containing 600mM sucrose) or by
10 homogenising 25mg of seed in 300µL Extraction Buffer using a Wiggenshauser D-130 Homogenizer. A further 750µL Extraction Buffer was added and the slurry in the mortar and transferred to a 2mL microfuge tube whereas the homogenizer tip was rinsed in 1mL Extraction Buffer and this volume was added to the homogenised seed. Samples were then centrifuged at 20,000 ×g for 5min; this left a pellet and aqueous supernatant which was overlaid by an
15 immiscible oily layer containing both intact and disrupted oil bodies as well as free TAG. The upper oil layer was gently pushed to the side of the tube, and the aqueous layer and pelleted material discarded. The oil layer was then re-suspended from the side of the tube by vortexing in Extraction Buffer and placed in a fresh 2mL microfuge tube. The final volume was made up to 0.5mL with Extraction Buffer.

20 Purified Oil Body preparations from Arabidopsis thaliana seeds and cross linking cysteine residues between the engineered oleosins

25 mg of *Arabidopsis* seed (of plants transformed as described in Example 5) was ground in 300 µl extraction buffer (10 mM Phosphate buffer, pH 7.5 containing 600 mM sucrose) using a Wiggenshauser D-130 Homogenizer. Seed was ground until crushed and the sample appeared
25 "creamy" and frothy as starch was released from the seeds. The homogenizer tip was rinsed in 1 ml buffer and this volume was added to the crushed seed. Samples were prepared up to this point in lots of 4, then centrifuged 14,000rpm for 5 mins. A thin gel loading tip was used to gently push the oil layer to the side of the tube, and the aqueous layer removed to a fresh tube. The oil layer was resuspended from the side of the tube using extraction buffer and placed in a
30 fresh 2 ml tube. The final volume was made up to 0.5 ml (as read on the side of the tube) with extraction buffer, samples were divided into two and oxidising agent (3mM GSSG) was added to

one tube and incubated at room temperature for 10 min. Oil body preparations were then added to an equal volume of 2 x gel loading buffer and boiled for 5min before loading on to a gel.

Samples were run either on pre-cast NuPAGE Novex 4-12% Bis-Tris Midi Gels (Invitrogen) on a Criterion gel rig system (Bio-Rad), or NuPAGE® Novex 12% Bis-Tris gradient Gel 1.0 mm, 15 well, cat# NP0343BOX, with NuPAGE® MES SDS Running Buffer (for Bis-Tris Gels only) (20X), cat# NP0002-02, or on hand-cast Tris-HCl gels. Gels were stained by SafeStain (Invitrogen) to show total protein loaded or blotted onto Nitrocellulose membrane using the iBlot system (Invitrogen). In each case, the negative control was a sample extracted from wild type Columbia seed and the positive control was the same extraction method (although grinding was by mortar and pestle) performed on wild type sesame seed. 10µl of each sample and the negative control were loaded onto the gel, and 5µl was used for the positive control.

Following blotting, the membrane was blocked in a solution of 12.5% skim milk powder in TBST (50 mM Tris pH 7.4, 100 mM NaCl, 0.2 % Tween) for at least 1.5 hours. The membrane was then washed in TBST 3 x 5 mins before incubating with primary antibody (anti-sesame) at 1/1000 in TBST for 1 hour at room temperature. Following 3 further TBST washes, incubation with secondary antibody (anti-rabbit) at 1/5000 was carried out for 1 hour at room temperature. The membrane underwent 3 further washes then the signal was developed using standard chemiluminescence protocol.

Figure 11 shows the accumulation of sesame seed oleosin units on the oil bodies under the control of the CaMV35S promoter. It can be seen that recombinant oleosin and polyoleosin was found to accumulate in the seeds of *Arabidopsis thaliana* and was correctly targeted to the oil bodies (Figure 11). In addition, it can be seen that in the presence of oxidising agent for 10 minutes the recombinant oleosins containing cysteines formed cross-links as evidenced by the appearance of oligomers and corresponding disappearance of the monomeric forms in these samples and not in the wild type or non oxidised transgenic oil bodies.

The effect of increasing the number of potential cross-linking sites in an oleosin peptide on *in planta* OB integrity and emulsion stability can be assessed as follows.

Quantitative Determination of OB integrity

Assessment of OB stability and integrity using either absorbance (OD₆₀₀), direct counting of AOBs using a hemocytometer, or visual evaluation of coalescence by microscopy proved to be highly variable and amongst other things was influenced by the: degree of pre-sampling

agitation; quantity of sample removed; time left under the microscope. To avoid this the applicants devised a simple method to quantify the amount of TAG released from the OBs into the surrounding media during a variety of treatments as a means of comparing integrity. Essentially equal quantities (based on FAMES-GC/MS estimation of TAG and Bradfords determination of protein) of OB preparations are made up to a total volume of 200 μ L using AOB
5 buffer (containing Proteinase K [PNK] when appropriate at a 1:1 ratio of PNK:total proteins in OB samples in a 250 μ L GC glass insert tubes and covered with a plastic cap. Following the treatment (elevated temperature or exposure to PNK) 15 μ L of fish oil (Vitamax®, Australia) is added to the sample and mixed by vortexing followed by centrifugation at 5,200g for 1min. The
10 addition of fish oil followed by vortexing enables any TAG that had leaked from the OBs to mix with the added fish oil and be floated by brief centrifugation. 4 μ L of the oil phase is sampled and subjected to fatty acid methyl esterification (FAME) and then analysed by GC-MS (Shimadzu model numbers, fitted with a 50mQC2/BPX70-0.25 GC capillary column (SGE) as described by Browse *et al.* (1986). In the absence of added fish oil the quantity of TAG that had
15 leaked from the OBs was too small to form a samplable visible layer even after centrifugation, in such a case the maximum volume would have been 6 μ L. The very different lipid profiles of fish oil and sesame oil enabled us to easily distinguish the leaked TAG from the added TAG.

Using the internal C15:0 and C17:0 standards the applicants can calculate the absolute amounts of C18:2 (the major lipid in sesame seed oil) recovered after treatment.

20 *Determination of OB integrity and emulsion stability at elevated temperature*

Oil in water emulsions are less stable at elevated temperatures; hence, it is of interest to investigate if modified oleosins with varying numbers in introduced cysteines influence OB and AOB integrity at elevated temperature. To achieve this the applicants determine the integrity (using the method described above) of OBs (containing different oleosins) in an phosphate buffer
25 (50mM Na-phosphate buffer pH8, 100mM NaCl) at 95°C. AOBs are heated for 2h. Integrity is determined as above.

The effect of higher ratios of crosslinked oleosin:TAG increase the stability of OBs in rumen fluid can be assessed as follows:

Determination of OB integrity in rumen fluid

One of the aims of disulfide was to provide some degree of protection from biohydrogenation by rumen microflora. Assessment of OB stability with rumen fluid can be assessed as follows. OBs are added to an equal volume (25 μ L) of rumen fluid. Samples are incubated at 39°C for 0, 15, 30, 60, 120 and 240min, at the end of the incubation an equal volume of loading buffer (Invitrogen) is added, mixed and heated at 70°C for 10min. 15 μ L of each sample/loading buffer mix is compared by SDS-PAGE/immunoblot. Integrity is determined as above.

Analysis of OB integrity in Proteinase K

To investigate the influence of modified oleosin in a controllable and repeatable highly degradative environment integrity is determined (using the method described above) of AOBs (containing different modified oleosins) after incubation in an phosphate buffer (50mM Na-phosphate buffer pH8, 100mM NaCl) containing 1:1 (g/g protein) Proteinase K (Invitrogen) at 37°C for 4h. While the maximum activity of Proteinase K is achieved below 65°C the lower temperature is used in order to reduce the influence of temperature on OB instability. Integrity is determined as above.

EXAMPLE 7: Production of oil bodies in the leaves of *Arabidopsis thaliana*

In order to produce oil bodies in vegetative tissue, it is necessary to produce triacylglycerol in such tissue (e.g. leaves).

Production of triacylglycerol in the vegetative portions of the plant

In most plants (including *Lolium perenne*) the majority of leaf lipids are attached to a glycerol backbone and exist as diacylglycerols. These are incorporated into lipid bi-layers where they function as membranes of multiple sub-cellular organelles or the as the membrane of the cell itself. The majority of lipid bilayer in the leaf is the chloroplast thylakoid membrane. A smaller amount of leaf lipid exists as epicuticular waxes and an even smaller percentage is present in the form of triacylglycerol (TAG).

Most plants synthesise and store TAG in developing embryos and pollen cells where it is subsequently utilised to provide catabolizable energy during germination and pollen tube growth. Dicotyledonous plants can accumulate up to approximately 60% of their seed weight as TAG. Ordinarily, this level is considerably lower in the monocotyledonous seeds where the main form of energy storage is carbohydrates (e.g., starch)The only committed step in TAG biosynthesis is

the last one, i.e., the addition of a third fatty acid to an existing diacylglycerol, thus generating TAG. In plants this step is performed by one of three enzymes including: acyl CoA:diacylglycerol acyltransferase (DGAT1), an unrelated acyl CoA:diacylglycerol acyl transferase (DGAT2), and phospholipid:diacylglycerol acyltransferase (Zou *et al.*, 1999; 5 Bouvier-Navé *et al.*, 2000; Dahlqvist *et al.*, 2000; Lardizabal *et al.*, 2001). Over expression of the transcribed region of any of these genes in the vegetative portions of plants leads to the formation of TAG droplets in the cytoplasm of leaf cells, as demonstrated by the over expression of: Arabidopsis DGAT1 in tobacco by Bouvier-Navé *et al.*, (2000); Tung tree DGAT2 in yeast and tobacco by Shockey *et al.*, (2006); Arabidopsis PDAT in Arabidopsis by Stahl *et al.*, (2004). 10 Over-expression of Arabidopsis DGAT1 in some cases was demonstrated to increase the total lipid level but not necessarily by the accumulation of TAG, e.g. in *Lotus japonicus* hairy roots (Bryan *et al.*, 2004) and in *Lolium perenne* leaves (Cookson *et al.*, 2009).

To demonstrate the accumulation of TAG in the leaves of these plants you can compare the total quantity of lipid extract from leaves of these plants with those of untransformed plants or plants 15 transformed with the empty binary vector. Ensuring the plants are grown under the same environmental conditions and that the leaves sampled are physiologically equivalent. With the appropriate internal standards the quantification of the total lipid extract can be achieved using FAMES GC-MS analysis (as described by Winichayakul *et al.*, 2008 Delivery of grasses with high levels of unsaturated, protected fatty acids. Proceedings of the New Zealand Grassland 20 Association, 70:211-216.). Alternatively, the total lipids can be extracted using the Folsch method (Folsch *et al.*, 1957 J. Folsch, M. Lees and G.A. Slone-Stanley, A simple method for the determination of total lipid extraction and purification, Journal of Biological Chemistry 226 (1957), pp. 497-507.) and quantified using appropriate internal standards with a GC-MS fitted with a Restek (Restek Corp., Bellefonte, PA) RTX65TG column.

25 Leaves were sampled from plants over expressing the *A. thaliana* DGAT1 (S205A) and the sesame seed oleosin construct (either Oleo_0-0, or Oleo_1-1, or Oleo_1-3, or Oleo_3-1, or Oleo_3-3, SEQ ID NOs 11-20, Figures 1-5) and analysed by SDS-PAGE/immunoblot using the polyclonal anti-sesame seed oleosin antisera. It can be seen that recombinant oleosin was found to accumulate in the leaves of *Arabidopsis thaliana* leaves (Figure 12).

30 The simultaneous expression and accumulation of oleosin/modified oleosin protein in the same cell (for example leaf cell) will result in the production of triglyceride oil bodies encapsulated by a phospholipid monolayer embedded with oleosin; this has been demonstrated with un-modified oleosin in yeast (Ting *et al.*, 1997) and seeds (Abell *et al.*, 2004).

Oil body preparations from the leaves of transgenic Arabidopsis thaliana

Oil bodies can be extracted from the leaves of transgenic *Arabidopsis thaliana* expressing DGAT1 (S205A) and the sesame seed oleosin construct (either Oleo_0-0, or Oleo_1-1, or Oleo_1-3, or Oleo_3-1, or Oleo_3-3, SEQ ID NOs 11-20, Figures 1-5).

- 5 The effect of increasing the number of potential cross-linking sites in an oleosin peptide on the OBs of such plants can be assessed by measuring OB integrity and emulsion stability can as described in Example 6.

Design and construction of oleosins containing more than three cysteine residues in each hydrophilic arms

- 10 The ole-3,3 lines had substantial levels of elevated lipid levels in the form of TAGs when co-expressed with DGAT1 (S205A) while the lines containing ole-0,0 did not have elevated lipid levels above the DGAT1 over expressing control. The ole-1,1, ole-1,3 and ole-3,1 showed there was a correlation between the level of lipid accumulation in the leaf and the increase in the number of cysteines engineered into each arm (Table 3).
- 15 Table 3. Fatty acid composition (as %Dry Weight) of *Arabidopsis* leaves expressing either vector control, DGAT1 (S205A) alone, or DGAT1 (S205A) and different forms of oleosin (containing either no additional cysteines or up to 3 additional cysteines in each hydrophilic arm).

Fatty acid profile	Vector control	DGAT1 ALONE DGAT1SA #2	DGAT1 +OLE 0-0 (#11)	DGAT1 +OLE 1-1 (#9)	DGAT1 +OLE 1-3 (#5)	DGAT1 +OLE 3-1 (#18)	DGAT1 +OLE 3-3 (#47)
C16:0	0.55±0.035	0.55±0.001	0.54±0.014	0.57±0.001	0.68±0.042*	0.62±0.084	0.95±0.049*
C16:1	0.085±0.007	0.105±0.007	0.11±0.001	0.13±0.014	0.1±0.001	0.135±0.021	0.11±0.001
C16:3	0.34±0.021	0.41±0.028	0.42±0.007	0.48±0.028	0.51±0.035	0.55±0.071*	0.62±0.049*
C18:1	0.095±0.007	0.075±0.007	0.1±0.001	0.185±0.007	0.345±0.007	0.2±0.014*	0.61±0.014*
C18:2	0.55±0.014	0.46±0.035	0.56±0.014	0.77±0.049*	0.97±0.007*	0.79±0.113*	1.82±0.113*
C18:3	1.67±0.056	1.91±0.028	1.78±0.014	1.68±0.028	1.74±0.014	1.9±0.28	2.29±0.056*
C20:0	Not detected	Not detected	Not detected	Not detected	Not detected	Not detected	0.054±0.003

- 20 The correlation between the increase in total lipid (shown to be TAG) and the number of cysteines engineered into the hydrophilic domains indicated that the number of cysteines may be a way to influence the level of TAG desired. Consequently new constructs containing more than 3 cysteines per hydrophilic arm were designed. While it is not possible to put an infinite number of cysteines/hydrophilic arm; the limitations include:

- Length of the arms - if additional residues were added to make space for the cysteines then eventually the degree of hydrophobic domain interaction would be reduced since their ability to come into contact would be limited by their freedom to move on the OB.

- Maintaining the proportion of +, - and amphipathic residues - if the balance of these residues and distribution of these residues is altered dramatically it is likely that the hydrophilic arms would not actually interact with the surface of the OB and as such would not provide any protection against lipases or coalescence.

- Sulfur availability – increasing the number of cysteines per oleosin molecule may place the plant under nutritional stress if sulphur is limiting.

10 The original cysteine-oleosin was engineered to carry 3 relatively evenly spaced unpaired cysteines in each arm by replacing amino acids and predominantly those that could be predicted to be neutral or charged but not hydrophobic.

The oleosin presumably needs to have a certain level of negative charge and in the C-terminus this appears to be achieved by K (Lys), hence continuing the strategy of swapping charged or neutral residues with additional cysteines may result in poor stability in terms of preventing coalescence. Furthermore, in the N-terminal hydrophilic region there appears to be too few residues left between the engineered cysteines to enable further substitution of residues whilst maintaining the spacing and oscillation between positive and negatively charged amino acids. Hence, for both N- and C-termini added additional residues (cysteines) rather than substitute existing residues with cysteines. Alternatively, an oleosin with longer hydrophilic arms could have been used.

Two additional constructs (Ole-5, 6 and Ole-6,7) were also designed. These are not purposely unbalanced in terms of cysteine residues per arm but organised to attempt to give typically 4-5 residues between each cysteine. In fact to increase the cysteines to 6 in the N-terminal arm it was necessary to generate additional residues (as opposed to substitution of existing residues); this as achieved by replicate the first 6 residues from the Ole-3,3.

Rather than have completely new nucleotide sequences designed the triplet TGT to code for cysteine was added (where appropriate) to generate Ole_5,6. For additional glutamine residues the codon triplet GGA was used. For the additional N-terminal 6 residues on Ole_6-7 the N-terminus of Ole_3,3 was replicated and fused in frame.

Subcloning strategy was designed to be identical to initial cysteine oleosins, i.e., subcloned into oleoacceptor by NotI/NdeI. This is then recombined by LR reaction into pRSH1 (Winichayakul et al., 2008). Essentially places both Arabidopsis DGAT1 (S205A) and oleosin under their own CaMV35s promoters and OCS terminators. Both DGA1 and oleosin clones contain a UBQ10
5 intron.

NetGene2 was used to predict the splicing pattern of Ole_5,6 and Ole_6,7. Both were predicted to have only one donor and acceptor site on the direct strand (both were predicted to have a very high probability of recognition) and no sites on the complementary strand.

The data indicates that the oleosins containing 1,3 or 3,1 cysteines do accumulate detectable
10 levels of TAG but this is certainly less than the 3,3 cysteine oleosins (the 1,1 accumulated trace amounts while the 0, 0 did not). This suggests even more strongly that the 5,6 and 6,7 oleosins are likely to accumulate even more TAG than the 3,3 construct. The first data from the 5,6 and 6,7 constructs will be available soon.

Transformation of oleosins containing engineered cysteines and DGAT1 into wild type
15 *Arabidopsis thaliana*

Five disulfide-oleosin/DGAT1 (S205A) gene constructs and one control (construct containing DGAT1 (S205A) but not oleosin) were been transferred to the plant binary vector pRSh1 (Winichayakul et al., 2008) and transformed into wild type Arabidopsis thaliana using Agrobacterium-mediated transformation.

20 A modification of the traditional floral dip method was followed since it has been reported that floral dipping tends to damage developing siliques due to the presence of detergent in the inoculums (Martinez-Trujillo et al., 2004). Therefore, a drop by drop inoculation to every flower was carried out using a micropipette. The inoculation was repeated after one week to introduce the inoculum to the newly developed flowers. Seeds were collected when the siliques have dried
25 up, then cleaned and planted for screening of transformants.

Screening for transformants was performed by BASTA selection and homozygous transformants were selected using segregation ratio analysis for BASTA resistance.

Transformation of oleosins containing engineered cysteines and DGAT1 into wild type Trifolium repens

Transformation into *Trifolium repens* (white clover) was performed according to the procedure of Voisey *et al.*, (1994).

5 Seeds were weighed to provide approximately 400 – 500 cotyledons (ie. 200 – 250 seeds) for dissection (0.06gm = 100 seeds). In a centrifuge tube, seeds were rinsed with 70% ethanol for 1 minute. Surface sterilised in bleach (5% available chlorine) by shaking on a circular mixer for 15 minutes followed by four washes in sterile water. Seeds were imbibed overnight at 4degC.

10 The same constructs used to transform Arabidopsis (abover) were maintained in Agrobacterium strain GV3101 and inoculated into 25 mL of MGL broth (Table 4) containing spectinomycin at a concentration of 100mg/L. Cultures were grown overnight (16 hours) on a rotary shaker (200rpm) at 28oC. Bacterial cultures were harvested by centrifugation (3000xg, 10 minutes). The supernatant was removed and the cells resuspended in a 5mL solution of 10mM MgSO₄.

15 Cotyledons were dissected from seeds using a dissecting microscope. First, the seed coat and endosperm were removed. Cotyledons were separated from the radical with the scalpel by placing the blade between the cotyledons and slicing through the remaining stalk. Cotyledonary explants were harvested onto a sterile filter disk on CR7 media.

For transformation, a 3ul aliquot of Agrobacterium suspension was dispensed to each dissected cotyledon. Plates were sealed and cultured at 25degC under a 16 hour photoperiod. Following a
20 72 hour period of co-cultivation, transformed cotyledons were transferred to plates containing CR7 medium supplemented with ammonium glufosinate (2.5mg/L) and timentin (300mg/L) and returned to the culture room.

Following the regeneration of shoots, explants were transferred to CR5 medium supplemented with ammonium glufosinate (2.5mg/L) and timentin (300mg/L). Regenerating shoots are
25 subcultured three weekly to fresh CR5 media containing selection.

As root formation occurs, plantlets were transferred into tubs containing CR0 medium containing ammonium glufosinate selection. Large clumps of regenerants were divided to individual plantlets at this stage. Whole, rooted plants growing under selection were then potted into sterile peat plugs. Once established in peat plugs plants were then transfer to the greenhouse.

30 Table 4. Media compositions used for *Trifolium repens* transformation.

A.	CR#0		
	MS salts		
	B5 vitamins		
	sucrose		30 g/L
	pH5.8		(KOH)
	agar		8.0 g/L
	CR#5		
	MS salts		
	B5 vitamins		
	sucrose		30 g/L
	BA		0.1mg/L
	NAA		0.05mg/L
	pH 5.8		(KOH)
	agar		8.0g/L
B.	CR#7		
	MS salts		
	B5 vitamins		
	sucrose		30 g/L
	BA		1.0mg/L
	NAA		0.05mg/L
	pH 5.8		(KOH)
	agar		8.0g/L
C.	MGL		
	Mannitol		5.0g/L
	L glutamic acid	1.0g/L	
	KH ₂ PO ₄		250mg/L
	MgSO ₄		100mg/L
	NaCl		100mg/L
	Biotin		100mg/L
	Bactotryptone	5.0g/L	
	Yeast extract	2.5g/L	
	pH7.0		(NaOH)

FAMES GC/MS results showed the transgenic *Trifolium repens* (containing DGAT1 (S205A) and either Ole_3,3 or Ole_5,6 or Ole 6,7) had elevated total leaf lipid profiles compared to wild

type (Figure 17). There was a general correlation between the highest level of leaf lipid and the highest number of cysteines engineered into the oleosin.

FAMES GC/MS results showed the transgenic *Trifolium repens* (containing DGAT1 (S205A) and either Ole_3,3 or Ole_5,6 or Ole_6,7) had elevated C18:1 and C18:2 leaf lipid profiles compared to wild type as also seen in *Arabidopsis* (Figure 18). The highest level of leaf C18:1 and C18:2 was found in plants transformed with the oleosin containing the highest number of engineered cysteines.

Determination of oil body assembly in leaves (and seeds)

Further screening was conducted using immunoblot analysis (with an anti-sesame seed oleosin antibody, Scott et al., 2007) to determine the lines overexpressing the oleosin protein. Using this method, either oil bodies (OBs) were extracted from T2 seeds of putative transformants using sucrose density gradient or total protein was extracted from leaves in a denaturing/reducing buffer and proteins were separated in SDS-PAGE, transferred to nitrocellulose membrane, and challenged with an antibody raised against the sesame oleosin (Scott et al., 2007).

Crude oil body (OB) was extracted from 25 mg of seeds in 500 μ L OB buffer (10 mM Sodium phosphate, pH 7.5 containing 600 mM sucrose). After centrifugation at 13,000 x g, the aqueous layer was carefully suck out and the fat pad layer was resuspended in 200 μ L of OB buffer without disturbing the pellet at the bottom. 20 μ L of each OB extract was added with 4x loading dye and 10x reducing agent, heated up to 70°C for 5 min and loaded onto 4-12% polyacrylamide gel for immunoblot analysis. The blot was incubated in α -sesame oleosin antibody (1°Ab) at 1:750 dilution for one hour, and another one hour in secondary antibody (1:10,000).

Oleosin is naturally expressed in seeds and not in the leaves. However, since we have co-expressed DGAT1 with oleosin both under the control of CaMV35S promoters it could be anticipated that this would enable detectable levels of oleosin to accumulate in the leaves.

Leaves from transformed lines with high expression of recombinant oleosin in the seeds (identified by immunoblot analysis) were analyzed by immunoblot using antibodies raised against the sesame oleosin.

Table 5 below summarises the number of putative transformants generated and the number of plants expressing recombinant oleosin in the seed and leaf.

Table 5

Gene construct ID	Number of putative transformants (based on BASTA resistance)	Number of lines seeds were analysed by immunoblot (anti sesame seed oleosin antibody)	Number of lines with a positive immunoreactive band at the appropriate size in the seed extract	Number of lines with a positive immunoreactive band at the appropriate size in the leaf extract
pRSh1-DGAT1(S2 05A) control	8	N/A	N/A	
pRSh1-DGAT1(S2 05A)-Ole_0-0	14	8	7	3
pRSh1-DGAT1(S2 05A)-Ole_1-1	22	2	1	1
pRSh1-DGAT1(S2 05A)-Ole_1-3	20	0	0	1
pRSh1-DGAT1(S2 05A)-Ole_3-1	23	8	4	2
pRSh1-DGAT1(S2 05A)-Ole_3-3	54	22	16	5

It should be noted the level of recombinant oleosin that accumulated in the leaves was considerably lower than in the seeds. However, the proportion of individual lines accumulating detectable levels in both the leaves was much greater than when oleosin was expressed alone

(Roberts Lab, unpublished data) indicating that the co-expression of both DGAT1 and oleosin in the leaf has lead to the accumulation of higher levels of oleosin.

Analysis of leaves from transgenic plants co-expressing DGAT1 (S205A) and disulfide oleosins

The seeds from homozygous lines over expressing the oleosin protein in the seeds were germinated to allow growth of 2, 3, 4 or 5 weeks. Sufficient leaf material was harvested for FAMES GC-MS, as well as by GC-MS using a RTX 65-TG Restek column which enable the separation and identification of free fatty acids, diacylglycerides, wax esters, sterol esters and triacylglycerides without derivatization.

Preparation of material for FAMES-GC/MS analysis

10 10 mg of freeze-dried leaf powder was placed in a 13 x 100 mm screw-cap tube, 10 μ L of non methylated internal standard (C15:0 FA, 4 mg/mL dissolved in heptane) was added, To this mixture, 1 mL of the methanolic HCl reagent (1 mL of 3 M solution diluted to 1 M using dry methanol which had been treated with 5% 2,2-dimeethoxypropane as a water scavenger). The tube was then flushed with N₂ gas then sealed immediately with Teflon-lined cap, and heated at
15 80 °C for 1 h. After the tubes had cooled to room temperature, 10 μ L pre-methylated standard (4 mg/mL of C17:0 dissolved in heptane) was added. To this mixture, 0.6 mL heptane and 1.0 mL of 0.9% (w/v) NaCl was added, and mixed thoroughly by vortexing. Following centrifugation at 500 rpm for 1 min at room temperature, 100 μ L of the top layer (containing heptanes) was collected and transferred to a flat-bottom glass insert fitted into a brown vial for GC/MS analysis.

20 *FAMES GC/MS Analysis*

The FAMES GC/MS was analysed using the SGE capillary column BPX70 (50m x 0.22mm x 0.25 μ m). The condition of GC-MS was as follows: the temperature was programmed from 80 °C to 150 °C at 15 °C /min and then to 250 °C at 8 °C /min and held isothermal for 10 min. Samples were injected in a split mode; total flow of 28.4 mL/min; column flow of 0.82 mL/min;
25 and a purge flow of 3.0 mL/min. The pressure was kept at 150 kPa; ion source temperature was 200 °C and an interface temperature was kept at 260 °C. The target compounds were acquired by mass spectrometry in a scan mode starting at 50 m/z and ending at 350 m/z.

TAG and polar lipid extraction

30 TAG was extracted using a modified method of Ruiz-López et al., (2003). Briefly, for each TAG analysis, between 34-80 mg of freeze-dried leaf powder was placed into tared 13-mm screw cap

tube and weighed, 2.4 mL of 0.17 M NaCl in MeOH was added and mixed by vortexing. Following the addition of 4.8 mL heptane and 10 μ L of internal standard (C14:0, 10 μ g. μ L⁻¹), the suspension was mixed gently and incubated without shaking in 80 °C water bath for 2 h. After cooling to room temperature, the upper phase (containing lipids) was transferred to fresh screw-cap tube and evaporated to dryness under stream of N gas. Finally, the dried powder were resuspended in 100 μ L heptanes, mixed thoroughly then transferred to a flat-bottom glass insert fitted into a brown glass vial for TAG analysis.

TAG GC-MS analysis

TAG analysis was performed on a Hewlett Packard (HP) GC and Shimadzu Scientific Instruments Inc. MS (QP2010). All analyses were performed with a RESTEK capillary column MXT-65TG (65% diphenyl - 35% dimethyl polysiloxane, 30.0 m x 0.10 μ m thickness x 0.25 mm diameter) in Electron Impact (EI) ionization mode. Helium was used as the carrier gas. All samples were injected in splitless mode, in 1.0 μ L aliquots, and a column flow of 1.2 mL.min⁻¹. The gas chromatograph was programmed from 200 to 370 °C at 15 °C.min⁻¹ and kept isothermal at 370 °C for 15 min. The sample injector port temperature was maintained at 350 °C, column oven temperature at 200 °C, with a pressure of 131.1 kPa and a purge flow of 3.0 mL.min⁻¹. The mass spectrometric conditions were as follows: ion source temperature was held at 260 °C during the GC-MS runs, the mass spectra were obtained at ionization voltage of 70 eV at an emission current of 60 μ A and an interface temperature of 350 °C. Acquisition mode was by scanning at a speed of 5000, 0.25 sec per scan. Chromatograph peaks with mass to charge ratio of 45 m/z to 1090m/z were collected starting at 9 min and ending at 25 min.

EXAMPLE 8: Further oleosins, caloleosins and steroleosins engineered to contain additional cysteine residues in the N- and C- terminal hydrophilic arms

The applicants have used the same strategy as for sesame seed oleosin, accession number AAD42942, (i.e., substituting charged residues predicted to be on the surface of OBs with cysteines) to engineer cysteines into the N- and C- terminal hydrophilic arms of oleosins caoleosins and steroleosins. In some cases it has been possible to substitute only negatively charged amino acids (Glutamic acid and Aspartic acid) that are relatively evenly spaced. In the case of the sesame oleosin AAD42942 it was necessary to sometimes compromise on the charge substitution. It should be noted in the examples below that two caleosins (AAB71227 and AAF13743) contain two endogenous cysteines in their C-terminal arm. These are left unaltered in the engineering.

To determine the position of the amino acid substitution each protein was aligned with the sesame oleosin (AAD42942) in the original form as well as the forms containing 1 or 3 cysteines per hydrophilic arm (i.e., ole_0,0; ole_1,1; ole_3,1; ole_1,3; ole_3,3). The potential glutamic acids and aspartic acids in N-terminus or C-terminus of each of the hydrophilic arms (determined by hydrophobicity plots) were then highlighted with grey boxes, as were the relevant lysine, arginine and glutamine residues (which were all successfully altered in the sesame oleosin (AAD42942). The mutation of these residues to cysteine were then considered along with their spacing with each other. The final substitutions are then shown with the original peptide sequence and the engineered sequence only. In this case only 3 cysteines were engineered into each arm, however, the number could have been greater or fewer. An alternative approach would have been to work with each protein in isolation and simply begin by identifying the hydrophilic regions by hydrophobicity plot then begin the process of substitution with the most appropriate charged amino acid.

Table 6 below shows additional oleosin and caoleosins that the applicants have modified to introduce cysteines in the hydrophilic portions.

Table 6.

Protein Type	Plant Source	Accession Number	SEQ ID NO
oleosin	<i>Brassica oleraceae</i> (pollen oleosin)	CAA65272.1	90
oleosin	Maize	NP_001147032.1	91
oleosin	Rice	AAL40177.1	92
caoleosin	Sesame	AAF13743	93
caoleosin	Soybean	AAB71227	94
caoleosin	Maize	NP_001151906	95
steroleosin	Sesame	AAL13315	96
steroleosin	Brassica napus	ACG69522	97
steroleosin	Maize	NP_001152614.1	98

Table 7 below references the SEQ ID NO in the modified oleosins

Protein Type	Plant Source	Accession Number	SEQ ID NO
oleosin	<i>Brassica oleraceae</i> (pollen oleosin)	X96409	99
oleosin	Maize	NP_001147032.1	100
oleosin	Rice	AAL40177.1	101
caoleosin	Sesame	AAF13743	102
caoleosin	Soybean	AAB71227	103
caoleosin	Maize	NP_001151906	104
steroleosin	Sesame	AAL13315	105
steroleosin	<i>Brassica napus</i>	ACG69522	106
steroleosin	Maize	NP_001152614.1	107

The modified sequence can be expressed as described in the examples above to produce oil bodies, emulsions, transgenic host cells, plants etc, and to test the properties of each.

It is not the intention to limit the scope of the invention to the abovementioned examples only.

As would be appreciated by a skilled person in the art, many variations are possible without

5 departing from the scope of the invention.

REFERENCES

- Abell *et al.*, (2004). Plant J., 37: 461-70.
- Altschul *et al.*, (1997) Nucleic Acids Res. 25: 3389-3402,
- Andrianov *et al.*, (2010). Plant Biotechnol J. 8(3):277-87.
- 5 Ausubel *et al.*, (1987) Current Protocols in Molecular Biology, Greene Publishing Bairoch and
Bucher (1994), Nucleic Acids Res. 22, 3583
- Bao *et al.*, (2000) Plant J. 22(1):39-50.
- Birch (1997) Ann Rev Plant Phys Plant Mol Biol, 48, 297
- Bolton and McCarthy (1962) PNAS 84:1390
- 10 Bowie *et al.*, (1990). Science 247, 1306
- Bouvier-Nave *et al.*, (2000) *Eur. J. Biochem.* 267, 85-96.
- Bryan *et al.*, (2007) Modification of fatty acid biosynthesis. United States Patent 20070118927.
- Capuano *et al.*, (2007). Biotechnol Adv. 25:203-206.
- Notredame *et al.*, (2000). J. Mol. Biol. 302: 205-217
- 15 Chen *et al.*, (1999). Plant Cell Physiol., 40:1079-1086. Chiang *et al.*, (2005). J Agric Food Chem
53:4799-804.
- Chiang *et al.*, (2007). Protein Expr Purif. 52:14-8.
- Cookson *et al.*, (2009). Improvements in and relating to oil production. PCT/NZ2008/000085
WO/2008/130248
- 20 Dahlqvist *et al.*, (2000). Proc Natl Acad Sci U S A. 97, 6487-6492.
- Deckers *et al.*, (2003). United States patent US 6582710
- Demeyer and Doreau, (1999). Proc Nutr Soc. 58(3):593-607.
- Deutscher (1990) Ed, Methods in Enzymology, Vol. 182, Guide to Protein Purification
- Draper *et al.*, 1988, Plant Genetic Transformation and Gene Expression. A Laboratory Manual,
25 Blackwell Sci. Pub. Oxford, p. 365
- Falquet *et al.*, 2002, Nucleic Acids Res. 30, 235
- Feng and Doolittle; 1987, J. Mol. Evol. 25, 351
- Firkins *et al.*, (2006). J Dairy Sci. 89 Suppl 1:E31-51. Review. Greenspan.
- Frandsen *et al.*, (2001). Physiologia Plantarum, 112:301-307.

- Frohman (1993). *Methods Enzymol.* 218: 340-56
- Galun and Breiman (1997). *Transgenic Plants*. Imperial College Press, London
- Gelvin *et al.*, 1993, *Plant Molecular Biol. Manual*. Kluwer Acad. Pub. Dordrecht
- Giesen *et al.*, *Nucleic Acids Res.* 1998 Nov 1;26(21):5004-6
- 5 Halford & Hardie (1998). *Plant Mol Biol.* 37:735-48. Review
- Harada *et al.*, (2002). OLEOSIN/PHOSPHOLIPID COMPLEX AND PROCESS FOR
PRODUCING THE SAME. World Patent WO/2002/026788.
- Hellens *et al.*, (2000). *Plant Mol Biol* 42: 819-32
- Hellens *et al.*, (2005). *Plant Meth* 1: 13
- 10 Herrera-Estrella *et al.*, (1993). *Nature* 303, 209
- Hofmann *et al.*, (1999). *Nucleic Acids Res.* 27, 215
- Hou *et al.*, (2003). *J Dairy Sci*; 86: 424-8.
- Huang (1992). *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 43:177-200.
- Huang, X. (1994) *Computer Applications in the Biosciences* 10, 227-235
- 15 Jeanmougin *et al.*, (1998) *Trends Biochem. Sci.* 23, 403-5.
- Jenkins and Bridges (2007). *Eur. J. Lipid Sci. Technol.* 109:778-789.
- Jenkins and McGuire (2006). *J Dairy Sci.* 89(4):1302-10. Review.
- Kaup *et al.*, (2002) *Plant Physiol.* 129(4):1616-26.
- Kyte and Doolittle (1982). *J. Mol. Biol.* 157:105-132
- 20 Lanfranco L. (2003). *Riv Biol.* 96(1):31-54.
- Lardizabal *et al.*, (2001). *J.B.C.* 276, 38862-38869.
- Leprince *et al.*, (1998). *Planta* 204 109-119.
- Lin and Tzen. (2004). *Plant Physiology and Biochemistry.* 42:601-608.
- Lock and Bauman (2004). *Lipids.* 39(12):1197-206. Review.
- 25 Loer and Herman (1993). *Plant Physiol.* 101(3):993-998.
- Mayer and Fowler (1985). *J Cell Biol.* 100(3):965-73.
- Mekhedov *et al.*, (2000). *Plant Physiol.* 122(2):389-402).
- Mullis *et al.*, Eds. 1994 *The Polymerase Chain Reaction*, Birkhauser
- Murphy (1993). *Prog. Lipid Res.* 32:247-280.

- Needleman and Wunsch, (1970) J. Mol. Biol. 48, 443-453
- Nielsen *et al.*, Science. (1991) 254(5037):1497-500
- Ohlrogge and Jaworski (1997). Annu Rev Plant Physiol Plant Mol Biol. 48:109-136.
- Papapostolou and Howorka (2009). Mol Biosyst. 5(7):723-32. Review.
- 5 Peng (2004). Development and applications of artificial sesame oil body. Ph.D. Dissertation. National ChunHsing University Graduate Institute of Biotechnology. Taichung, Taiwan.
- Peng *et al.*, (2006) . Stability enhancement of native and artificial oil bodies by genipin crosslink. Taiwan patent I250466.
- 10 Peng *et al.*, (2004). J Biotechnol 2004; 111: 51-7.
- Potrykus and Spangenburg (1995). Gene Transfer to Plants. Springer-Verlag, Berlin
- Roberts *et al.*, (2008). The Open Biotechnology Journal 2:13-21.
- Roux *et al.*, (2004). J Agric Food Chem. 52(16):5245-9.
- Scott *et al.*, (2007). Polyoleosins. WO2007045019.
- 15 Saha *et al.*, (2006). Plant Physiol. 141(4):1533-43.
- Sambrook *et al.*, Eds, 1987, Molecular Cloning, A Laboratory Manual, 2nd Ed. Cold Spring Harbor Press
- Sarmiento *et al.*, (1997). Plant J. 11(4):783-96.
- Schrott (1995) In: Gene Transfer to Plants (Potrykus, T., Spangenberg. Eds) Springer Verlag.
- 20 Berline, pp. 325-336
- Shimada *et al.*, (2008). Plant J. 55(5):798-809.
- Shockey *et al.*, (2006). Plant Cell., 18,2294-2313.
- Siloto *et al.*, (2006). Plant Cell. 18(8):1961-74.
- Slack *et al.*, (1980). Biochem J. 190(3):551-561.
- 25 Slocombe *et al.*, (2009). Plant Biotechnol J. 7(7):694-703.
- Stahl *et al.*, (2004). Plant Physiology, 135: 1324-1335.
- Stewart *et al.*, (1969), In: Solid-Phase Peptide Synthesis, WH Freeman Co, San Francisco California
- Thompson *et al.*, (1994) Nucleic Acids Research, 22:4673-4680

- Ting *et al.*, (1997). J Biol Chem., 272: 3699-3706.
- Tadege *et al.*, (2005). Trends Plant Sci. 10(5):229-35.
- Triglia *et al.*, 1998, Nucleic Acids Res 16, 8186
- Tzen *et al.*, (1992). J. Biol. Chem. 267: 15626-34
- 5 Tzen *et al.*, (2003). Adv Plant Physiol., 6: 93-104.
- Tzen *et al.*, (1997). J Biochem. 121(4):762-8.
- Voisey *et al.*, (1994). Plant Cell Reports 13: 309 314.
- Winichayakul *et al.*, (2008). Proc. NZGA, 70:211-216
- Xu *et al.*, (2005). Plant Cell. 17(11):3094-110.
- 10 Zou *et al.*, (1999). Plant J. 19, 645-653.
- Zou *et al.*, (2008). Plant Biotech. J. 6(8):799-818.

SUMMARY OF SEQUENCE LISTING

SEQ ID NO:	Type	SPECIES	COMMENTS
1	polynucleotide	artificial	Oleosin disulfide 0,0 nucleotide sequence, as cloned into pET29b using NdeI and XhoI restriction sites (adds N-terminal S•tag thrombin cleavage site and C-terminal His tag).
2	polynucleotide	artificial	Oleosin disulfide 1,1 nucleotide sequence, as cloned into pET29b using NdeI and XhoI restriction sites (adds N-terminal S•tag thrombin cleavage site and C-terminal His tag).
3	polynucleotide	artificial	Oleosin disulfide 1,3 nucleotide sequence, as cloned into pET29b using NdeI and XhoI restriction sites (adds N-terminal S•tag thrombin cleavage site and C-terminal His tag).
4	polynucleotide	artificial	Oleosin disulfide 3,1 nucleotide sequence, as cloned into pET29b using NdeI and XhoI restriction sites (adds N-terminal S•tag thrombin cleavage site and C-terminal His tag).
5	polynucleotide	artificial	Oleosin disulfide 3,3 nucleotide sequence, as cloned into pET29b using NdeI and XhoI restriction sites (adds N-terminal S•tag thrombin cleavage site and C-terminal His tag).
6	Polypeptide	Artificial	Oleosin disulfide 0,0 peptide sequence, as cloned into pET29b using NdeI and XhoI restriction sites (adds N-terminal S•tag thrombin cleavage site and C-

			terminal His tag).
7	Polypeptide	Artificial	Oleosin disulfide 1,1 peptide sequence, as cloned into pET29b using NdeI and XhoI restriction sites (adds N-terminal S-tag thrombin cleavage site and C-terminal His tag).
8	Polypeptide	Artificial	Oleosin disulfide 1,3 peptide sequence, as cloned into pET29b using NdeI and XhoI restriction sites (adds N-terminal S-tag thrombin cleavage site and C-terminal His tag).
9	Polypeptide	Artificial	Oleosin disulfide 3,1 peptide sequence, as cloned into pET29b using NdeI and XhoI restriction sites (adds N-terminal S-tag thrombin cleavage site and C-terminal His tag).
10	Polypeptide	Artificial	Oleosin disulfide 3,3 peptide sequence, as cloned into pET29b using NdeI and XhoI restriction sites (adds N-terminal S-tag thrombin cleavage site and C-terminal His tag).
11	Polynucleotide	Artificial	(Nucleotide sequence of Oleosin disulfide 0,0 including Kozac sequence and UBQ10 intron, as transformed into <i>Arabidopsis thaliana</i> under the control of the CaMV35s promoter.)
12	Polynucleotide	Artificial	Nucleotide sequence of Oleosin disulfide 1,1 including Kozac sequence and UBQ10 intron, as transformed into <i>Arabidopsis thaliana</i> under the control of the CaMV35s promoter.
13	Polynucleotide	Artificial	(Nucleotide sequence of Oleosin disulfide 1,3 including Kozac sequence

			and UBQ10 intron, as transformed into <i>Arabidopsis thaliana</i> under the control of the CaMV35s promoter.)
14	Polynucleotide	Artificial	Nucleotide sequence of Oleosin disulfide 3,1 including Kozac sequence and UBQ10 intron, as transformed into <i>Arabidopsis thaliana</i> under the control of the CaMV35s promoter.
15	Polynucleotide	Artificial	Nucleotide sequence of Oleosin disulfide 3,3 including Kozac sequence and UBQ10 intron, as transformed into <i>Arabidopsis thaliana</i> under the control of the CaMV35s promoter.
16	Polypeptide	Artificial	Peptide sequence of Oleosin disulfide 0,0, as transformed into <i>Arabidopsis thaliana</i> under the control of the CaMV35s promoter.
17	Polypeptide	Artificial	Peptide sequence of Oleosin disulfide 1,1, as transformed into <i>Arabidopsis thaliana</i> under the control of the CaMV35s promoter.)
18	Polypeptide	Artificial	Peptide sequence of Oleosin disulfide 1,3, as transformed into <i>Arabidopsis thaliana</i> under the control of the CaMV35s promoter.
19	Polypeptide	Artificial	Peptide sequence of Oleosin disulfide 3,1, as transformed into <i>Arabidopsis thaliana</i> under the control of the CaMV35s promoter.
20	Polypeptide	Artificial	Peptide sequence of Oleosin disulfide 3,3, as transformed into <i>Arabidopsis thaliana</i> under the control of the

			CaMV35s promoter
21	Polynucleotide	Artificial	Nucleotide sequence of Oleosin disulfide 5,6 including Kozac sequence and UBQ10 intron, as transformed into <i>Arabidopsis thaliana</i> under the control of the CaMV35s promoter.
22	Polynucleotide	Artificial	Nucleotide sequence of Oleosin disulfide 6,7 including Kozac sequence and UBQ10 intron, as transformed into <i>Arabidopsis thaliana</i> under the control of the CaMV35s promoter.
23	Polypeptide	Artificial	Peptide sequence of Oleosin disulfide 5,6, as transformed into <i>Arabidopsis thaliana</i> under the control of the CaMV35s promoter
24	Polypeptide	Artificial	Peptide sequence of Oleosin disulfide 6,7, as transformed into <i>Arabidopsis thaliana</i> under the control of the CaMV35s promoter
25	Polynucleotide	Artificial	Oleoacceptor (contains OCS terminator, CAMV35S promoter, DGAT1 (S205A) from <i>Arabidopsis</i> and UBQ10 intron)
26	Polynucleotide	Artificial	Oleosin_0,0 and DGAT1 (S205A) in pRSH1
27	Polynucleotide	Artificial	Oleosin _1,1 and DGAT1 (S205A) in pRSH1
28	Polynucleotide	Artificial	Oleosin _1,3 and DGAT1 (S205A) in pRSH1
29	Polynucleotide	Artificial	Oleosin _3,1 and DGAT1 (S205A)

			in pRSH1
30	Polynucleotide	Artificial	Oleosin _3,3 and DGAT1 (S205A) in pRSH1
31	Polynucleotide	Artificial	Oleosin _5,6 and DGAT1 (S205A) in pRSH1
32	Polynucleotide	Artificial	Oleosin _6,7 and DGAT1 (S205A) in pRSH1
33	Polypeptide	Artificial	DGAT1 (S205A)
34	Polynucleotide	<i>S. indicum</i>	Oleosin - AF302807
35	Polypeptide	<i>S. indicum</i>	Oleosin - AAG23840
36	Polynucleotide	<i>S. indicum</i>	Oleosin - U97700
37	Polypeptide	<i>S. indicum</i>	Oleosin - AAB58402
38	Polynucleotide	<i>A. thaliana</i>	Oleosin - X62353
39	Polypeptide	<i>A. thaliana</i>	Oleosin - CAA44225
40	Polynucleotide	<i>A. thaliana</i>	Oleosin - BT023738
41	Polypeptide	<i>A. thaliana</i>	Oleosin - AAZ23930
42	Polynucleotide	<i>H. annuus</i>	Oleosin - X62352.1
43	Polypeptide	<i>H. annuus</i>	Oleosin - CAA44224.1
44	Polynucleotide	<i>B. napus</i>	Oleosin - X82020.1
45	Polypeptide	<i>B. napus</i>	Oleosin - CAA57545.1
46	Polynucleotide	<i>Z. mays</i>	Oleosin - NM_001153560.1
47	Polypeptide	<i>Z. mays</i>	Oleosin - NP_001147032.1

48	Polynucleotide	<i>O. sativa</i>	Oleosin - L76464
49	Polypeptide	<i>O. sativa</i>	Oleosin - AAL40177.1
50	Polynucleotide	<i>B. oleracea</i>	Oleosin - AF117126.1
51	Polypeptide	<i>B. oleracea</i>	Oleosin - AAD24547.1
52	Polynucleotide	<i>C. arabica</i>	Oleosin - AY928084.1
53	Polypeptide	<i>C. arabica</i>	Oleosin - AAY14574.1
54	Polynucleotide	<i>S. indicum</i>	Steroleosin - AF421889
55	Polypeptide	<i>S. indicum</i>	Steroleosin - AAL13315
56	Polynucleotide	<i>B. napus</i>	Steroleosin - EU678274
57	Polypeptide	<i>B. napus</i>	Steroleosin - ACG69522
58	Polynucleotide	<i>Z. mays</i>	Steroleosin - NM_001159142.1
59	Polypeptide	<i>Z. mays</i>	Steroleosin - NP_001152614.1
60	Polynucleotide	<i>B. napus</i>	Steroleosin - EF143915.1
61	Polypeptide	<i>B. napus</i>	Steroleosin - ABM30178.1
62	Polynucleotide	<i>S. indicum</i>	Caleosin - AF109921
63	Polypeptide	<i>S. indicum</i>	Caleosin - AAF13743
64	Polynucleotide	<i>G. max</i>	Caleosin - AF004809
65	Polypeptide	<i>G. max</i>	Caleosin - AAB71227
66	Polynucleotide	<i>Z. mays</i>	Caleosin - NM_001158434.1
67	Polypeptide	<i>Z. mays</i>	Caleosin - NP_001151906
68	Polynucleotide	<i>B. napus</i>	Caleosin - AY966447.1
69	Polypeptide	<i>B. napus</i>	Caleosin - AAY40837.1

70	Polynucleotide	<i>C. revoluta</i>	Caleosin - FJ455154.1
71	Polypeptide	<i>C. revoluta</i>	Caleosin - ACJ70083.1
72	Polynucleotide	<i>C. sativus</i>	Caleosin - EU232173.1
73	Polypeptide	<i>C. sativus</i>	Caleosin - ABY56103.1
74	Polynucleotide	<i>A. thaliana</i>	DGAT1 - NM_127503
75	Polypeptide	<i>A. thaliana</i>	DGAT1 - NP_179535
76	Polynucleotide	<i>T. majus</i>	DGAT1 - AY084052
77	Polypeptide	<i>T. majus</i>	DGAT1 - AAM03340
78	Polynucleotide	<i>Z. mays</i>	DGAT1 - EU039830.1
79	Polypeptide	<i>Z. mays</i>	DGAT1 - ABV91586.1
80	Polynucleotide	<i>A. thaliana</i>	DGAT2 - NM_115011
81	Polypeptide	<i>A. thaliana</i>	DGAT2 - NP_566952.1
82	Polynucleotide	<i>B. napus</i>	DGAT2 - FJ858270
83	Polypeptide	<i>B. napus</i>	DGAT2 - AC090187.1
84	Polynucleotide	<i>A. hypogaea</i>	DGAT3 (soluble DGAT) - AY875644
85	Polypeptide	<i>A. hypogaea</i>	DGAT3 (soluble DGAT) - AAX62735.1
86	Polynucleotide	<i>A. thaliana</i>	PDAT - NM_121367
87	Polypeptide	<i>A. thaliana</i>	PDAT - NP_196868.1
88	Polynucleotide	<i>R. communis</i>	PDAT - XM_002521304

89	Polypeptide	<i>R. communis</i>	PDAT - XP_002521350
90	Polypeptide	<i>B. oleracea</i>	Oleosin - CAA65272.1
91	Polypeptide	<i>Z. mays</i>	Oleosin - NP_001147032.1
92	Polypeptide	<i>O. sativa</i>	Oleosin - AAL40177.1
93	Polypeptide	<i>S. indicum</i>	Caleosin - AAF13743
94	Polypeptide	<i>G. Max</i>	Caleosin - AAB71227
95	Polypeptide	<i>Z. mays</i>	Caleosin - NP_001151906
96	Polypeptide	<i>S. indicum</i>	Steroleosin- AAL13315
97	Polypeptide	<i>Brassica napus</i>	steroleosin ACG69522
98	Polypeptide	<i>Z. mays</i>	steroleosin NP_001152614.1
99	Polypeptide	<i>Brassica oleracea</i>	Modified pollen oleosin – CAA65272.1
100	Polypeptide	<i>Zea mays</i>	Modified oleosin - NP_001147032.1
101	Polypeptide	<i>Oryza sativa</i>	Modified oleosin - AAL40177.1
102	Polypeptide	<i>S. indicum</i>	Modified caleosin - AAF13743
103	Polypeptide	<i>G. soja</i>	Modified caleosin - AAB71227
104	Polypeptide	<i>Z. mays</i>	Modified caleosin – NP_001151906
105	Polypeptide	<i>S. indicum</i>	Modified steroleosin- AAL13315
106	Polypeptide	<i>Brassica napus</i>	Modified steroleosin- ACG69522
107	Polypeptide	<i>Z. mays</i>	Modified steroleosin- NP_001152614.1

CLAIMS:

1. A polynucleotide encoding a modified oleosin including at least one artificially introduced cysteine, wherein the cysteine is introduced into at least one of:
 - a) the N-terminal hydrophilic region of the oleosin, and
 - b) the C-terminal hydrophilic region of the oleosin.
2. The polynucleotide of claim 1 in which the polynucleotide encodes a fusion protein including the modified oleosin fused to a protein of interest.
3. A genetic construct, expression construct, host cell, plant cell or plant, wherein the construct, cell or plant comprises the polynucleotide of any one of claims 1 and 2.
4. The host cell, plant cell or plant of claim 3 that is also genetically modified to express a triacylglycerol (TAG) synthesising enzyme.
5. The host cell of any one of claims 3 and 4 that is a host cell selected from a bacterial cell, a yeast cell, a fungal cell, an insect cell, an algal cell, and a plant cell.
6. A modified oleosin that:
 - a) is encoded by the polynucleotide of claim 1, or
 - b) includes at least one artificially introduced cysteine, wherein the cysteine is introduced into at least one of:
 - a) the N-terminal hydrophilic region of the oleosin, and
 - b) the C-terminal hydrophilic region of the oleosin.
7. A fusion protein comprising the modified oleosin of claim 6 and and a protein of interest.
8. An oil body, emulsion, composition, or composition with a carrier, wherein the oil body, emulsion or composition comprises at least one of:
 - a) the modified oleosin of claim 6, and
 - b) the fusion protein of claim 7.

9. An emulsion, composition, composition with carrier, plant or plant thereof, vegetative tissue of a plant, seed of a plant, animal feed, animal feed comprising a plant or tissue thereof, wherein the composition, plant, plant part, tissue, seed or animal feed comprises an oil body of claim 8.

10. A method for producing an oil body, the method comprising the step of combining

- a) at least two of the modified oleosins of claim 6,
- b) triacylglycerol, and
- c) phospholipid.

11. The method of claim 10 in which the components of a), b) and c) are combined as specified in at least one of:

- i) within a host cell;
- ii) within a host cell in which the modified oleosins are expressed;
- iii) within the host cell of i) or ii) that is also genetically modified to express a triacylglycerol (TAG) synthesising enzyme;
- iv) within the host cell of any one of i) – iii) that forms part of an organism;
- v) within the host cell of any one of i) – iii) that forms part of a plant;
- vi) within the host cell of any one of i) – iii) that forms part of a plant that accumulates about 50% to about to about 400% more lipid than does a suitable control plant; and
- vii) *in vitro*;

wherein the method optionally includes the additional step of purifying the oil bodies from the cell or organism.

12. A method of producing a plant that accumulates more oil than a suitable control plant the method comprising providing a plant comprising the polynucleotide of claim 1 that expresses a modified oleosin encoded by the polynucleotide.

13. The method of claim of claim 12 wherein the plant is also transformed with a

polynucleotide encoding a TAG synthesising enzyme to express the TAG synthesising enzyme and thus synthesise TAG.

14. The method of any one of claims 12 to 13 wherein the plant accumulates about 50% to about 400% more lipid than does a suitable control plant.

15. The method of claim any one of claims 12 to 14 wherein the plant is processed into at least one of:

- i) an animal feed; and
- ii) a biofuel feed stock.

16. A method for producing an oil body in a host cell, the method comprising:

- a) introducing into a host cell at least one polynucleotide of any one of claims 1 and 2; and
- b) culturing the host cell in order to express the modified oleosin.

17. A method for producing an oil body in a host cell, the method comprising:

- a) introducing into a host cell at least one polynucleotide of any one of claims 1 and 2 and a nucleic acid molecule encoding a TAG synthesizing enzyme ; and
- b) culturing the host cell in order to express the modified oleosin and the TAG synthesizing enzyme.

18. The method of any one of claims 16 and 17 wherein the host cell is processed into an oil fraction.

19. The method of any one of claims 18 wherein the oil is processed into a fuel, oleochemical or nutritional or cosmetic oil, a polyunsaturated fatty acid (PUFA) or a combination thereof.

Figure 1

CaMV 35S
~~~~~  
T CGA CGA ATT AAT TCC AAT CCC ACA AAA ATC  
CaMV 35S  
~~~~~  
TGA GCT TAA CAG CAC AGT TGC TCC TCT CAG AGC AGA ATC GGG TAT TCA ACA CCC TCA TAT
CaMV 35S
~~~~~  
CAA CTA CTA CGT TGT GTA TAA CGG TCC ACA TGC CGG TAT ATA CGA TGA CTG GGG TTG TAC  
CaMV 35S  
~~~~~  
AAA GGC GGC AAC AAA CGG CGT TCC CGG AGT TGC ACA CAA GAA ATT TGC CAC TAT TAC AGA
CaMV 35S
~~~~~  
GGC AAG AGC AGC AGC TGA CGC GTA CAC AAC AAG TCA GCA AAC AGA CAG GTT GAA CTT CAT  
CaMV 35S  
~~~~~  
CCC CAA AGG AGA AGC TCA ACT CAA GCC CAA GAG CTT TGC TAA GGC CCT AAC AAG CCC ACC
CaMV 35S
~~~~~  
AAA GCA AAA AGC CCA CTG GCT CAC GCT AGG AAC CAA AAG GCC CAG CAG TGA TCC AGC CCC  
CaMV 35S  
~~~~~  
AAA AGA GAT CTC CTT TGC CCC GGA GAT TAC AAT GGA CGA TTT CCT CTA TCT TTA CGA TCT
CaMV 35S
~~~~~  
AGG AAG GAA GTT CGA AGG TGA AGG TGA CGA CAC TAT GTT CAC CAC TGA TAA TGA GAA GGT  
CaMV 35S  
~~~~~  
TAG CCT CTT CAA TTT CAG AAA GAA TGC TGA CCC ACA GAT GGT TAG AGA GGC CTA CGC AGC
CaMV 35S
~~~~~  
AGG TCT CAT CAA GAC GAT CTA CCC GAG TAA CAA TCT CCA GGA GAT CAA ATA CCT TCC CAA  
CaMV 35S  
~~~~~  
GAA GGT TAA AGA TGC AGT CAA AAG ATT CAG GAC TAA TTG CAT CAA GAA CAC AGA GAA AGA
CaMV 35S
~~~~~  
CAT ATT TCT CAA GAT CAG AAG TAC TAT TCC AGT ATG GAC GAT TCA AGG CTT GCT TCA TAA  
CaMV 35S  
~~~~~  
ACC AAG GCA AGT AAT AGA GAT TGG AGT CTC TAA AAA GGT AGT TCC TAC TGA ATC TAA GGC
CaMV 35S
~~~~~  
CAT GCA TGG AGT CTA AGA TTC AAA TCG AGG ATC TAA CAG AAC TCG CCG TGA AGA CTG GCG  
CaMV 35S  
~~~~~  
AAC AGT TCA TAC AGA GTC TTT TAC GAC TCA ATG ACA AGA AGA AAA TCT TCG TCA ACA TGG
CaMV 35S
~~~~~  
TGG AGC ACG ACA CTC TGG TCT ACT CCA AAA ATG TCA AAG ATA CAG TCT CAG AAG ACC AAA  
CaMV 35S

GGG CTA TTG AGA CTT TTC AAC AAA GGA TAA TTT CGG GAA ACC TCC TCG GAT TCC ATT GCC  
CaMV 35S

CAG CTA TCT GTC ACT TCA TCG AAA GGA CAG TAG AAA AGG AAG GTG GCT CCT ACA AAT GCC  
CaMV 35S

ATC ATT GCG ATA AAG GAA AGG CTA TCA TTC AAG ATC TCT CTG CCG ACA GTG GTC CCA AAG  
CaMV 35S

ATG GAC CCC CAC CCA CGA GGA GCA TCG TGG AAA AAG AAG ACG TTC CAA CCA CGT CTT CAA  
CaMV 35S

AGC AAG TGG ATT GAT GTG ACA TCT CCA CTG ACG TAA GGG ATG ACG CAC AAT CCC ACT ATC  
CaMV 35S

CTT CGC AAG ACC CTT CCT CTA TAT AAG GAA GTT CAT TTC ATT TGG AGA GGA CAC GCT CGA  
attB1

GGA ATT CGG TAC CCC ATC ACA AGT TTG TAC AAA AAA GCA GGC TGC GGC CGC TTG CTC CCT  
5' Oleo\_0-0

NcoI

M A E H Y G Q Q Q Q T R A P H L Q  
TAA AAA AAA CCA TGG CAG AGC ATT ACG GAC AAC AGC AAC AGA CTA GAG CAC CTC ATC TTC  
5' Oleo\_0-0

UBQ10

L Q P R A Q R V V K A A T A V T  
AGC TTC AAC CTA GAG CAC AGA GAG TTG TGA AGG CTG CTA CTG CTG TTA CTG TAA ATT TCT  
UBQ10

GTG TTC CTT ATT CTC TCA AAA TCT TCG ATT TTG TTT TCG TTC GAT CCC AAT TTC GTA TAT  
UBQ10

GTT CTT TGG TTT AGA TTC TGT TAA TCT TAG ATC GAA GAC GAT TTT CTG GGT TTG ATC GTT  
UBQ10

AGA TAT CAT CTT AAT TCT CGA TTA GGG TTT CAT AGA TAT CAT CCG ATT TGT TCA AAT AAT  
UBQ10

TTG AGT TTT GTC GAA TAA TTA CTC TTC GAT TTG TGA TTT CTA TCT AGA TCT GGT GTT AGT  
3'Oleo\_0-0

UBQ10

A G  
TTC TAG TTT GTG CGA TCG AAT TTG TCG ATT AAT CTG AGT TTT TCT GAT TAA CAG GCT GGA  
3' Oleo\_0-0

G S L L V L S G L T L A G T V I A L T I  
GGA TCT CTT CTT GTT CTC TCT GGA CTT ACT CTC GCT GGA ACT GTT ATC GCT CTC ACT ATC  
3' Oleo\_0-0

~~~~~  
A T P L L V I F S P V L V P A V I T I F
GCT ACA CCT CTT CTC GTT ATC TTC TCT CCT GTT CTC GTT CCT GCT GTG ATC ACT ATC TTC
3' Oleo_0-0
~~~~~

L L G A G F L A S G G F G V A A L S V L  
CTT CTC GGA GCT GGA TTT CTT GCT TCT GGT GGA TTT GGA GTT GCT GCT CTC TCT GTT CTC  
3' Oleo\_0-0  
~~~~~

S W I Y R Y L T G K H P P G A D Q L E S
TCT TGG ATC TAC AGA TAC CTC ACT GGA AAA CAT CCT CCA GGT GCT GAT CAA CTT GAG TCT
3' Oleo_0-0
~~~~~

A K T K L A S K A R E M K D R A E Q F S  
GCT AAG ACT AAG CTC GCT TCT AAG GCT AGA GAG ATG AAG GAT AGA GCA GAG CAA TTC TCT  
3' Oleo\_0-0 OCS terminator  
~~~~~

Q Q P V A G S Q T S * *
CAA CAG CCT GTT GCT GGA TCT CAG ACT TCT TAA TGA ACA TAT GGT CCT GCT TTA ATG AGA
OCS terminator
~~~~~

TAT GCG AGA CGC CTA TGA TCG CAT GAT ATT TGC TTT CAA TTC TGT TGT GCA CGT TGT AAA  
OCS terminator  
~~~~~

AAA CCT GAG CAT GTG TAG CTC AGA TCC TTA CCG CCG GTT TCG GTT CAT TCT AAT GAA TAT
OCS terminator
~~~~~

ATC ACC CGT TAC TAT CGT ATT TTT ATG AAT AAT ATT CTC CGT TCA ATT TAC TGA TTG TAC  
OCS terminator  
~~~~~

CCT ACT ACT TAT ATG TAC AAT ATT AAA ATG AAA ACA ATA TAT TGT GCT GAA TAG GTT TAT
OCS terminator
~~~~~

AGC GAC ATC TAT GAT AGA GCG CCA CAA TAA CAA ACA ATT GCG TTT TAT TAT TAC AAA TCC  
OCS terminator  
~~~~~

AAT TTT AAA AAA AGC GGC AGA ACC GGT CAA ACC TAA AAG ACT GAT TAC ATA AAT CTT ATT
OCS terminator
~~~~~

CAA ATT TCA AAA GGC CCC AGG GGC TAG TAT CTA CGA CAC ACC GAG CGG CGA ACT AAT AAC  
OCS terminator  
~~~~~

GTT CAC TGA AGG GAA CTC CGG TTC CCC GCC GGC GCG CAT GGG TGA GAT TCC TTG AAG TTG
OCS terminator
~~~~~

AGT ATT GGC CGT CCG CTC TAC CGA AAG TTA CGG GCA CCA TTC AAC CCG GTC CAG CAC GGC  
OCS terminator  
~~~~~

GGC CGG GTA ACC GAC TTG CTG CCC CGA GAA TTA TGC AGC ATT TTT TTG GTG TAT GTG GGC
OCS terminator
~~~~~

CCC AAA TGA AGT GCA GGT CAA ACC TTG ACA GTG ACG ACA AAT CGT TGG GCG GGT CCA GGG  
OCS terminator  
~~~~~

CGA ATT TTG CGA CAA CAT GTC GAG GCT CAG CAG GAC CTG CAG GCA TGC AAG CTA GCT TAC
~~~~~

OCS terminator

CaMV35S

TAG TGA TGC ATA TTC TAT AGT GTC ACC TAA ATC TTC GAC GAA TTA ATT CCA ATC CCA CAA

CaMV35S

AAA TCT GAG CTT AAC AGC ACA GTT GCT CCT CTC AGA GCA GAA TCG GGT ATT CAA CAC CCT

CaMV35S

CAT ATC AAC TAC TAC GTT GTG TAT AAC GGT CCA CAT GCC GGT ATA TAC GAT GAC TGG GGT

CaMV35S

TGT ACA AAG GCG GCA ACA AAC GGC GTT CCC GGA GTT GCA CAC AAG AAA TTT GCC ACT ATT

CaMV35S

ACA GAG GCA AGA GCA GCA GCT GAC GCG TAC ACA ACA AGT CAG CAA ACA GAC AGG TTG AAC

CaMV35S

TTC ATC CCC AAA GGA GAA GCT CAA CTC AAG CCC AAG AGC TTT GCT AAG GCC CTA ACA AGC

CaMV35S

CCA CCA AAG CAA AAA GCC CAC TGG CTC ACG CTA GGA ACC AAA AGG CCC AGC AGT GAT CCA

CaMV35S

GCC CCA AAA GAG ATC TCC TTT GCC CCG GAG ATT ACA ATG GAC GAT TTC CTC TAT CTT TAC

CaMV35S

GAT CTA GGA AGG AAG TTC GAA GGT GAA GGT GAC GAC ACT ATG TTC ACC ACT GAT AAT GAG

CaMV35S

AAG GTT AGC CTC TTC AAT TTC AGA AAG AAT GCT GAC CCA CAG ATG GTT AGA GAG GCC TAC

CaMV35S

GCA GCA GGT CTC ATC AAG ACG ATC TAC CCG AGT AAC AAT CTC CAG GAG ATC AAA TAC CTT

CaMV35S

CCC AAG AAG GTT AAA GAT GCA GTC AAA AGA TTC AGG ACT AAT TGC ATC AAG AAC ACA GAG

CaMV35S

AAA GAC ATA TTT CTC AAG ATC AGA AGT ACT ATT CCA GTA TGG ACG ATT CAA GGC TTG CTT

CaMV35S

CAT AAA CCA AGG CAA GTA ATA GAG ATT GGA GTC TCT AAA AAG GTA GTT CCT ACT GAA TCT

CaMV35S

AAG GCC ATG CAT GGA GTC TAA GAT TCA AAT CGA GGA TCT AAC AGA ACT CGC CGT GAA GAC

CaMV35S

TGG CGA ACA GTT CAT ACA GAG TCT TTT ACG ACT CAA TGA CAA GAA GAA AAT CTT CGT CAA

CaMV35S

CAT GGT GGA GCA CGA CAC TCT GGT CTA CTC CAA AAA TGT CAA AGA TAC AGT CTC AGA AGA

CaMV35S

CCA AAG GGC TAT TGA GAC TTT TCA ACA AAG GAT AAT TTC GGG AAA CCT CCT CGG ATT CCA  
CaMV35S

TTG CCC AGC TAT CTG TCA CTT CAT CGA AAG GAC AGT AGA AAA GGA AGG TGG CTC CTA CAA  
CaMV35S

ATG CCA TCA TTG CGA TAA AGG AAA GGC TAT CAT TCA AGA TCT CTC TGC CGA CAG TGG TCC  
CaMV35S

CAA AGA TGG ACC CCC ACC CAC GAG GAG CAT CGT GGA AAA AGA AGA CGT TCC AAC CAC GTC  
CaMV35S

TTC AAA GCA AGT GGA TTG ATG TGA CAT CTC CAC TGA CGT AAG GGA TGA CGC ACA ATC CCA  
CaMV35S

CTA TCC TTC GCA AGA CCC TTC CTC TAT ATA AGG AAG TTC ATT TCA TTT GGA GAG GAC ACG  
5' DGAT1 (S205A)

NcoI

M A I L D S A G V T T V  
GGA TCC TTG CTC CGT TAA AAA AAA CCA TGG CTA TCC TCG ATT CTG CTG GTG TTA CTA CTG  
5' DGAT1 (S205A)

T E N G G G E F V D L D R L R R R K S R  
TGA CTG AGA ATG GTG GTG GAG AGT TCG TTG ATC TCG ATA GAC TCA GAA GAA GAA AGT CTA  
5' DGAT1 (S205A)

UBQ10

S  
GAT CTG TAA ATT TCT GTG TTC CTT ATT CTC TCA AAA TCT TCG ATT TTG TTT TCG TTC GAT  
UBQ10

CCC AAT TTC GTA TAT GTT CTT TGG TTT AGA TTC TGT TAA TCT TAG ATC GAA GAC GAT TTT  
UBQ10

CTG GGT TTG ATC GTT AGA TAT CAT CTT AAT TCT CGA TTA GGG TTT CAT AGA TAT CAT CCG  
UBQ10

ATT TGT TCA AAT AAT TTG AGT TTT GTC GAA TAA TTA CTC TTC GAT TTG TGA TTT CTA TCT  
UBQ10

AGA TCT GGT GTT AGT TTC TAG TTT GTG CGA TCG AAT TTG TCG ATT AAT CTG AGT TTT TCT  
UBQ10

3' DGAT1 (S205A)

D S S N G L L L S G S D N N S P S  
GAT TAA CAG GAT TCT TCT AAC GGA CTT CTC CTC TCT GGA TCT GAT AAC AAC TCT CCT TCT  
3' DGAT1 (S205A)

D D V G A P A D V R D R I D S V V N D D  
GAT GAT GTT GGT GCT CCT GCT GAT GTG AGA GAT AGA ATC GAT TCT GTG GTG AAC GAT GAT  
3' DGAT1 (S205A)

~~~~~  
A Q G T A N L A G D N N G G G D N N G G
GCT CAA GGT ACT GCT AAC CTC GCT GGT GAT AAT AAC GGT GGA GGT GAT AAC AAT GGT GGA
3' DGAT1 (S205A)
~~~~~

G R G G G E G R G N A D A T F T Y R P S  
GGA AGA GGT GGA GGT GAA GGT AGA GGA AAC GCT GAT GCT ACT TTC ACT TAC AGA CCA TCT  
3' DGAT1 (S205A)  
~~~~~

V P A H R R A R E S P L S S D A I F K Q
GTG CCT GCT CAT AGA AGA GCT AGA GAG TCT CCT CTC TCT TCT GAT GCT ATC TTC AAG CAG
3' DGAT1 (S205A)
~~~~~

S H A G L F N L C V V V L I A V N S R L  
TCT CAC GCT GGA CTT TTC AAC CTC TGT GTG GTG GTT CTT ATC GCT GTG AAC TCT AGA CTC  
3' DGAT1 (S205A)  
~~~~~

I I E N L M K Y G W L I R T D F W F S S
ATC ATC GAG AAC CTC ATG AAG TAC GGA TGG CTC ATC AGA ACT GAT TTC TGG TTC TCT TCT
3' DGAT1 (S205A)
~~~~~

R S L R D W P L F M C C I S L S I F P L  
AGA TCT CTC AGA GAT TGG CCT CTT TTC ATG TGC TGC ATC TCA CTC TCA ATC TTC CCT CTC  
3' DGAT1 (S205A)  
~~~~~

A A F T V E K L V L Q K Y I A E P V V I
GCT GCT TTT ACT GTT GAG AAG CTC GTG CTC CAG AAG TAT ATC GCT GAA CCT GTG GTG ATC
3' DGAT1 (S205A)
~~~~~

F L H I I I T M T E V L Y P V Y V T L R  
TTC CTC CAC ATC ATC ATC ACT ATG ACT GAG GTT CTC TAC CCT GTT TAC GTG ACT CTC AGA  
3' DGAT1 (S205A)  
~~~~~

C D S A F L S G V T L M L L T C I V W L
TGC GAT TCT GCT TTC CTC TCT GGT GTT ACT CTT ATG CTC CTC ACT TGC ATT GTG TGG CTT
3' DGAT1 (S205A)
~~~~~

K L V S Y A H T S Y D I R S L A N A A D  
AAG CTC GTG TCT TAC GCT CAC ACT TCT TAC GAT ATC AGA TCT CTC GCT AAC GCT GCT GAT  
3' DGAT1 (S205A)  
~~~~~

K A N P E V S Y Y V S L K S L A Y F M V
AAG GCT AAC CCT GAA GTG TCT TAC TAC GTG TCT CTC AAG TCT CTC GCT TAC TTC ATG GTT
3' DGAT1 (S205A)
~~~~~

A P T L C Y Q P S Y P R S A C I R K G W  
GCT CCT ACA CTT TGT TAC CAG CCA TCT TAC CCT AGA TCT GCT TGC ATT AGA AAG GGA TGG  
3' DGAT1 (S205A)  
~~~~~

V A R Q F A K L V I F T G F M G F I I E
GTG GCA AGA CAA TTC GCT AAG TTG GTG ATC TTC ACT GGA TTC ATG GGA TTC ATC ATC GAG
3' DGAT1 (S205A)
~~~~~

Q Y I N P I V R N S K H P L K G D L L Y  
CAG TAC ATC AAC CCT ATT GTG AGA AAC TCT AAG CAC CCT CTC AAG GGT GAT CTT CTC TAC  
~~~~~

3' DGAT1 (S205A)

~~~~~  
A I E R V L K L S V P N L Y V W L C M F  
GCT ATC GAG AGA GTT CTT AAG CTC TCT GTG CCT AAC CTT TAT GTG TGG CTC TGC ATG TTC

3' DGAT1 (S205A)

~~~~~  
Y C F F H L W L N I L A E L L C F G D R
TAC TGT TTC TTC CAC CTC TGG CTT AAC ATC CTT GCT GAG TTG CTT TGC TTC GGA GAT AGA

3' DGAT1 (S205A)

~~~~~  
E F Y K D W W N A K S V G D Y W R M W N  
GAG TTC TAC AAG GAT TGG TGG AAC GCT AAG TCT GTT GGA GAT TAT TGG AGA ATG TGG AAC

3' DGAT1 (S205A)

~~~~~  
M P V H K W M V R H I Y F P C L R S K I
ATG CCT GTG CAT AAG TGG ATG GTG CGT CAC ATC TAC TTC CCT TGC CTC AGA TCT AAG ATC

3' DGAT1 (S205A)

~~~~~  
P K T L A I I I A F L V S A V F H E L C  
CCT AAG ACT CTC GCT ATC ATT ATC GCT TTC CTC GTG TCT GCT GTT TTC CAT GAG TTG TGT

3' DGAT1 (S205A)

~~~~~  
I A V P C R L F K L W A F L G I M F Q V
ATC GCT GTT CCT TGC AGA CTT TTC AAG CTT TGG GCT TTC CTC GGA ATC ATG TTC CAG GTT

3' DGAT1 (S205A)

~~~~~  
P L V F I T N Y L Q E R F G S T V G N M  
CCA CTC GTG TTC ATC ACT AAC TAC CTC CAA GAG AGA TTC GGA TCT ACT GTT GGA AAC ATG

3' DGAT1 (S205A)

~~~~~  
I F W F I F C I F G Q P M C V L L Y Y H
ATT TTC TGG TTC ATT TTC ATC TTC GGA CAG CCT ATG TGC GTT CTC CTC TAC TAC CAC

3' DGAT1 (S205A)

~~~~~  
D L M N R K G S M S \* \*  
GAT CTC ATG AAC AGA AAG GGA TCT ATG TCT TAA TGA AGG ATC CAC CCA GCT TTC TTG TAC

attB2

ocs

~~~~~  
AAA GTG GTG ATG GGT TCG AAA TCG ATA AGC TTG GAT CCT CTA GAG TCC TGC TTT AAT GAG

attB2

ocs

~~~~~  
ATA TGC GAG ACG CCT ATG ATC GCA TGA TAT TTG CTT TCA ATT CTG TTG TGC ACG TTG TAA

ocs

~~~~~  
AAA ACC TGA GCA TGT GTA GCT CAG ATC CTT ACC GCC GGT TTC GGT TCA TTC TAA TGA ATA

ocs

~~~~~  
TAT CAC CCG TTA CTA TCG TAT TTT TAT GAA TAA TAT TCT CCG TTC AAT TTA CTG ATT GTA

ocs

~~~~~  
CCC TAC TAC TTA TAT GTA CAA TAT TAA AAT GAA AAC AAT ATA TTG TGC TGA ATA GGT TTA

ocs

~~~~~  
TAG CGA CAT CTA TGA TAG AGC GCC ACA ATA ACA AAC AAT TGC GTT TTA TTA TTA CAA ATC

OCS

~~~~~  
CAA TTT TAA AAA AAG CGG CAG AAC CGG TCA AAC CTA AAA GAC TGA TTA CAT AAA TCT TAT

OCS

~~~~~  
TCA AAT TTC AAA AGG CCC CAG GGG CTA GTA TCT ACG ACA CAC CGA GCG GCG AAC TAA TAA

OCS

~~~~~  
CGT TCA CTG AAG GGA ACT CCG GTT CCC CGC CGG CGC GCA TGG GTG AGA TTC CTT GAA GTT

OCS

~~~~~  
GAG TAT TGG CCG TCC GCT CTA CCG AAA GTT ACG GGC ACC ATT CAA CCC GGT CCA GCA CGG

OCS

~~~~~  
CGG CCG GGT AAC CGA CTT GCT GCC CCG AGA ATT ATG CAG CAT TTT TTT GGT GTA TGT GGG

OCS

~~~~~  
CCC CAA ATG AAG TGC AGG TCA AAC CTT GAC AGT GAC GAC AAA TCG TTG GGC GGG TCC AGG

OCS

~~~~~  
GCG AAT TTT GCG ACA ACA TGT CGA GGC TCA GCA GGA CCT GCA GGC ATG CAA GCT AGC TTA

OCS

~~~~~  
CTA GTG ATG CAT ATT CTA TAG TGT CAC CTA AAT CT



Figure 2

```
CaMV 35S
T CGA CGA ATT AAT TCC AAT CCC ACA AAA ATC
CaMV 35S
TGA GCT TAA CAG CAC AGT TGC TCC TCT CAG AGC AGA ATC GGG TAT TCA ACA CCC TCA TAT
CaMV 35S
CAA CTA CTA CGT TGT GTA TAA CGG TCC ACA TGC CGG TAT ATA CGA TGA CTG GGG TTG TAC
CaMV 35S
AAA GGC GGC AAC AAA CGG CGT TCC CGG AGT TGC ACA CAA GAA ATT TGC CAC TAT TAC AGA
CaMV 35S
GGC AAG AGC AGC AGC TGA CGC GTA CAC AAC AAG TCA GCA AAC AGA CAG GTT GAA CTT CAT
CaMV 35S
CCC CAA AGG AGA AGC TCA ACT CAA GCC CAA GAG CTT TGC TAA GGC CCT AAC AAG CCC ACC
CaMV 35S
AAA GCA AAA AGC CCA CTG GCT CAC GCT AGG AAC CAA AAG GCC CAG CAG TGA TCC AGC CCC
CaMV 35S
AAA AGA GAT CTC CTT TGC CCC GGA GAT TAC AAT GGA CGA TTT CCT CTA TCT TTA CGA TCT
CaMV 35S
AGG AAG GAA GTT CGA AGG TGA AGG TGA CGA CAC TAT GTT CAC CAC TGA TAA TGA GAA GGT
CaMV 35S
TAG CCT CTT CAA TTT CAG AAA GAA TGC TGA CCC ACA GAT GGT TAG AGA GGC CTA CGC AGC
CaMV 35S
AGG TCT CAT CAA GAC GAT CTA CCC GAG TAA CAA TCT CCA GGA GAT CAA ATA CCT TCC CAA
CaMV 35S
GAA GGT TAA AGA TGC AGT CAA AAG ATT CAG GAC TAA TTG CAT CAA GAA CAC AGA GAA AGA
CaMV 35S
CAT ATT TCT CAA GAT CAG AAG TAC TAT TCC AGT ATG GAC GAT TCA AGG CTT GCT TCA TAA
CaMV 35S
ACC AAG GCA AGT AAT AGA GAT TGG AGT CTC TAA AAA GGT AGT TCC TAC TGA ATC TAA GGC
CaMV 35S
GAT GCA TGG AGT CTA AGA TTC AAA TCG AGG ATC TAA CAG AAC TCG CCG TGA AGA CTG GCG
CaMV 35S
AAC AGT TCA TAC AGA GTC TTT TAC GAC TCA ATG ACA AGA AGA AAA TCT TCG TCA ACA TGG
CaMV 35S
TGG AGC ACG ACA CTC TGG TCT ACT CCA AAA ATG TCA AAG ATA CAG TCT CAG AAG ACC AAA
CaMV 35S
```

GGG CTA TTG AGA CTT TTC AAC AAA GGA TAA TTT CGG GAA ACC TCC TCG GAT TCC ATT GCC  
CaMV 35S

CAG CTA TCT GTC ACT TCA TCG AAA GGA CAG TAG AAA AGG AAG GTG GCT CCT ACA AAT GCC  
CaMV 35S

ATC ATT GCG ATA AAG GAA AGG CTA TCA TTC AAG ATC TCT CTG CCG ACA GTG GTC CCA AAG  
CaMV 35S

ATG GAC CCC CAC CCA CGA GGA GCA TCG TGG AAA AAG AAG ACG TTC CAA CCA CGT CTT CAA  
CaMV 35S

AGC AAG TGG ATT GAT GTG ACA TCT CCA CTG ACG TAA GGG ATG ACG CAC AAT CCC ACT ATC  
CaMV 35S

CTT CGC AAG ACC CTT CCT CTA TAT AAG GAA GTT CAT TTC ATT TGG AGA GGA CAC GCT CGA  
attB1

GGA ATT CGG TAC CCC ATC ACA AGT TTG TAC AAA AAA GCA GGC TGC GGC CGC TTG CTC CCT  
5' Oleo\_1-1

NcoI

M A C H Y G Q Q Q Q T R A P H L Q  
TAA AAA AAA CCA TGG CAT GTC ATT ACG GAC AAC AGC AAC AGA CTA GAG CAC CTC ATC TTC  
5' Oleo\_1-1

UBQ10

L Q P R A Q R V V K A A T A V T  
AGC TTC AAC CTA GAG CAC AGA GAG TTG TGA AGG CTG CTA CTG CTG TTA CTG TAA ATT TCT  
UBQ10

GTG TTC CTT ATT CTC TCA AAA TCT TCG ATT TTG TTT TCG TTC GAT CCC AAT TTC GTA TAT  
UBQ10

GTT CTT TGG TTT AGA TTC TGT TAA TCT TAG ATC GAA GAC GAT TTT CTG GGT TTG  
UBQ10

AGA TAT CAT CTT AAT TCT CGA TTA GGG TTT CAT AGA TAT CAT CCG ATT TGT TCA AAT AAT  
UBQ10

TTG AGT TTT GTC GAA TAA TTA CTC TTC GAT TTG TGA TTT CTA TCT AGA TCT GGT GTT AGT  
3'Oleo\_1-1

UBQ10

A G  
TTC TAG TTT GTG CGA TCG AAT TTG TCG ATT AAT CTG AGT TTT TCT GAT TAA CAG GCT GGA  
3' Oleo\_1-1

G S L L V L S G L T L A G T V I A L T I  
GGA TCT CTT CTT GTT CTC TCT GGA CTT ACT CTC GCT GGA ACT GTT ATC GCT CTC ACT ATC  
3' Oleo\_1-1

~~~~~  
A T P L L V I F S P V L V P A V I T I F
GCT ACA CCT CTT CTC GTT ATC TTC TCT CCT GTT CTC GTT CCT GCT GTG ATC ACT ATC TTC
3' Oleo_1-1
~~~~~

L L G A G F L A S G G F G V A A L S V L  
CTT CTC GGA GCT GGA TTT CTT GCT TCT GGT GGA TTT GGA GTT GCT GCT CTC TCT GTT CTC  
3' Oleo\_1-1  
~~~~~

S W I Y R Y L T G K H P P G A D Q L E S
TCT TGG ATC TAC AGA TAC CTC ACT GGA AAA CAT CCT CCA GGT GCT GAT CAA CTT GAG TCT
3' Oleo_1-1
~~~~~

A K T K L A S K A R E M K D R A E Q F S  
GCT AAG ACT AAG CTC GCT TCT AAG GCT AGA GAG ATG AAG GAT AGA GCA GAG CAA TTC TCT  
3' Oleo\_1-1 OCS terminator  
~~~~~

C Q P V A G S Q T S * *
TGT CAG CCT GTT GCT GGA TCT CAG ACT TCT TAA TGA ACA TAT GGT CCT GCT TTA ATG AGA
OCS terminator
~~~~~

TAT GCG AGA CGC CTA TGA TCG CAT GAT ATT TGC TTT CAA TTC TGT TGT GCA CGT TGT AAA  
OCS terminator  
~~~~~

AAA CCT GAG CAT GTG TAG CTC AGA TCC TTA CCG CCG GTT TCG GTT CAT TCT AAT CAA TAT
OCS terminator
~~~~~

ATC ACC CGT TAC TAT CGT ATT TTT ATG AAT AAT ATT CTC CGT TCA ATT TAC TGA TTG TAC  
OCS terminator  
~~~~~

CCT ACT ACT TAT ATG TAC AAT ATT AAA ATG AAA ACA ATA TAT TGT GCT GAA TAG GTT TAT
OCS terminator
~~~~~

AGC GAC ATC TAT GAT AGA GCG CCA CAA TAA CAA ACA ATT GCG TTT TAT TAT TAC AAA TCC  
OCS terminator  
~~~~~

AAT TTT AAA AAA AGC GGC AGA ACC GGT CAA ACC TAA AAG ACT GAT TAC ATA AAT CTT ATT
OCS terminator
~~~~~

CAA ATT TCA AAA GGC CCC AGG GGC TAG TAT CTA CGA CAC ACC GAG CGG CGA ACT AAT AAC  
OCS terminator  
~~~~~

GTT CAC TGA AGG GAA CTC CGG TTC CCC GCC GGC GCG CAT GGG TGA GAT TCC TTG AAG TTG
OCS terminator
~~~~~

AGT ATT GGC CGT CCG CTC TAC CGA AAG TTA CGG GCA CCA TTC AAC CCG GTC CAG CAC GGC  
OCS terminator  
~~~~~

GGC CGG GTA ACC GAC TTG CTG CCC CGA GAA TTA TGC AGC ATT TTT TTG GTG TAT GTG GGC
OCS terminator
~~~~~

CCC AAA TGA AGT GCA GGT CAA ACC TTG ACA GTG ACG ACA AAT CGT TGG GCG GGT CCA GGC  
OCS terminator  
~~~~~

CGA ATT TTG CGA CAA CAT GTC GAG GCT CAG CAG GAC CTG CAG GCA TGC AAG CTA GCT TAC
~~~~~

OCS terminator

CaMV35S

TAG TGA TGC ATA TTC TAT AGT GTC ACC TAA ATC TTC GAC GAA TTA ATT CCA ATC GCA CAA

CaMV35S

AAA TCT GAG CTT AAC AGC ACA GTT GCT CCT CTC AGA GCA GAA TCG GGT ATT CAA CAC CCT

CaMV35S

CAT ATC AAC TAC TAC GTT GTG TAT AAC GGT CCA CAT GCC GGT ATA TAC GAT GAC TGG GGT

CaMV35S

TGT ACA AAG GCG GCA ACA AAC GGC GTT CCC GGA GTT GCA CAC AAG AAA TTT GCC ACT ATT

CaMV35S

ACA GAG GCA AGA GCA GCA GCT GAC GCG TAC ACA ACA AGT CAG CAA ACA GAC AGG TTG AAC

CaMV35S

TTC ATC CCC AAA GGA GAA GCT CAA CTC AAG CCC AAG AGC TTT GCT AAG GCC CTA ACA AGC

CaMV35S

CCA CCA AAG CAA AAA GCC CAC TGG CTC ACG CTA GGA ACC AAA AGG CCC AGC AGT GAT CCA

CaMV35S

GCC CCA AAA GAG ATC TCC TTT GCC CCG GAG ATT ACA ATG GAC GAT TTC CTC TAT CTT TAC

CaMV35S

GAT CTA GGA AGG AAG TTC GAA GGT GAA GGT GAC GAC ACT ATG TTC ACC ACT GAT AAT GAG

CaMV35S

AAG GTT AGC CTC TTC AAT TTC AGA AAG AAT GCT GAC CCA CAG ATG GTT AGA GAG GCC TAC

CaMV35S

GCA GCA GGT CTC ATC AAG ACG ATC TAC CCG AGT AAC AAT CTC CAG GAG ATC AAA TAC CTT

CaMV35S

CCC AAG AAG GTT AAA GAT GCA GTC AAA AGA TTC AGG ACT AAT TGC ATC AAG AAC ACA GAG

CaMV35S

AAA GAC ATA TTT CTC AAG ATC AGA AGT ACT ATT CCA GTA TGG ACG ATT CAA GGC TTG CTT

CaMV35S

CAT AAA CCA AGG CAA GTA ATA GAG ATT GGA GTC TCT AAA AAG GTA GTT CCT ACT GAA TCT

CaMV35S

AAG GCC ATG CAT GGA GTC TAA GAT TCA AAT CGA GGA TCT AAC AGA ACT CGC CGT GAA GAC

CaMV35S

TGG CGA ACA GTT CAT ACA GAG TCT TTT ACG ACT CAA TGA CAA GAA GAA AAT CTT CGT CAA

CaMV35S

CAT GGT GGA GCA CGA CAC TCT GGT CTA CTC CAA AAA TGT CAA AGA TAC AGT CTC AGA AGA

CaMV35S

CCA AAG GGC TAT TGA GAC TTT TCA ACA AAG GAT AAT TTC GGG AAA CCT CCT CGG ATT CCA  
CaMV35S

TTG CCC AGC TAT CTG TCA CTT CAT CGA AAG GAC AGT AGA AAA GGA AGG TGG CTC CTA CAA  
CaMV35S

ATG CCA TCA TTG CGA TAA AGG AAA GGC TAT CAT TCA AGA TCT CTC TGC CGA CAG TGG TCC  
CaMV35S

CAA AGA TGG ACC CCC ACC CAC GAG GAG CAT CGT GGA AAA AGA AGA CGT TCC AAC CAC GTC  
CaMV35S

TTC AAA GCA AGT GGA TTG ATG TGA CAT CTC CAC TGA CGT AAG GGA TGA CGC ACA ATC CCA  
CaMV35S

CTA TCC TTC GCA AGA CCC TTC CTC TAT ATA AGG AAG TTC ATT TCA TTT GGA GAG GAC ACG  
5' DGAT1 (S205A)

NcoI

M A I L D S A G V T T V  
GGA TCC TTG CTC CGT TAA AAA AAA CCA TGG CTA TCC TCG ATT CTG CTG GTG TTA CTA CTG  
5' DGAT1. (S205A)

T E N G G E F V D L D R L R R R K S R  
TGA CTG AGA ATG GTG GTG GAG AGT TCG TTG ATC TCG ATA GAC TCA GAA GAA GAA AGT CTA  
5' DGAT1 (S205A)

UBQ10

S  
GAT CTG TAA ATT TCT GTG TTC CTT ATT CTC TCA AAA TCT TCG ATT TTG TTT TCG TTC GAT  
UBQ10

CCC AAT TTC GTA TAT GTT CTT TGG TTT AGA TTC TGT TAA TCT TAG ATC GAA GAC GAT TTT  
UBQ10

CTG GGT TTG ATC GTT AGA TAT CAT CTT AAT TCT CGA TTA GGG TTT CAT AGA TAT CAT CCG  
UBQ10

ATT TGT TCA AAT AAT TTG AGT TTT GTC GAA TAA TTA CTC TTC GAT TTG TGA TTT CTA TCT  
UBQ10

AGA TCT GGT GTT AGT TTC TAG TTT GTG CGA TCG AAT TTG TCG ATT AAT CTG AGT TTT TCT  
UBQ10

3' DGAT1 (S205A)

D S S N G L L L S G S D N N S P S  
GAT TAA CAG GAT TCT TCT AAC GGA CTT CTC CTC TCT GGA TCT GAT AAC AAC TCT CCT TCT  
3' DGAT1 (S205A)

D D V G A P A D V R D R I D S V V N D D  
GAT GAT GTT GGT GCT CCT GCT GAT GTG AGA GAT AGA ATC GAT TCT GTG GTG AAC GAT GAT  
3' DGAT1 (S205A)

~~~~~  
A Q G T A N L A G D N N G G G D N N G G
GCT CAA GGT ACT GCT AAC CTC GCT GGT GAT AAT AAC GGT GGA GGT GAT AAC AAT GGT GGA
3' DGAT1 (S205A)
~~~~~

G R G G G E G R G N A D A T F T Y R P S  
GGA AGA GGT GGA GGT GAA GGT AGA GGA AAC GCT GAT GCT ACT TTC ACT TAC AGA CCA TCT  
3' DGAT1 (S205A)  
~~~~~

V P A H R R A R E S P L S S D A I F K Q
GTG CCT GCT CAT AGA AGA GCT AGA GAG TCT CCT CTC TCT TCT GAT GCT ATC TTC AAG CAG
3' DGAT1 (S205A)
~~~~~

S H A G L F N L C V V V L I A V N S R L  
TCT CAC GCT GGA CTT TTC AAC CTC TGT GTG GTG GTT CTT ATC GCT GTG AAC TCT AGA CTC  
3' DGAT1 (S205A)  
~~~~~

I I E N L M K Y G W L I R T D F W F S S
ATC ATC GAG AAC CTC ATG AAG TAC GGA TGG CTC ATC AGA ACT GAT TTC TGG TTC TCT TCT
3' DGAT1 (S205A)
~~~~~

R S L R D W P L F M C C I S L S I F P L  
AGA TCT CTC AGA GAT TGG CCT CTT TTC ATG TGC TGC ATC TCA CTC TCA ATC TTC CCT CTC  
3' DGAT1 (S205A)  
~~~~~

A A F T V E K L V L Q K Y I A E P V V I
GCT GCT TTT ACT GTT GAG AAG CTC GTG CTC CAG AAG TAT ATC GCT GAA CCT GTG GTG ATC
3' DGAT1 (S205A)
~~~~~

F L H I I I T M T E V L Y P V Y V T L R  
TTC CTC CAC ATC ATC ATC ACT ATG ACT GAG GTT CTC TAC CCT GTT TAC GTG ACT CTC AGA  
3' DGAT1 (S205A)  
~~~~~

C D S A F L S G V T L M L L T C I V W L
TGC GAT TCT GCT TTC CTC TCT GGT GTT ACT CTT ATG CTC CTC ACT TGC ATT GTG TGG CTT
3' DGAT1 (S205A)
~~~~~

K L V S Y A H T S Y D I R S L A N A A D  
AAG CTC GTG TCT TAC GCT CAC ACT TCT TAC GAT ATC AGA TCT CTC GCT AAC GCT GCT GAT  
3' DGAT1 (S205A)  
~~~~~

K A N P E V S Y Y V S L K S L A Y F M V
AAG GCT AAC CCT GAA GTG TCT TAC TAC GTG TCT CTC AAG TCT CTC GCT TAC TTC ATG GTT
3' DGAT1 (S205A)
~~~~~

A P T L C Y Q P S Y P R S A C I R K G W  
GCT CCT ACA CTT TGT TAC CAG CCA TCT TAC CCT AGA TCT GCT TGC ATT AGA AAG GGA TGG  
3' DGAT1 (S205A)  
~~~~~

V A R Q F A K L V I F T G F M G F I I E
GTG GCA AGA CAA TTC GCT AAG TTG GTG ATC TTC ACT GGA TTC ATG GGA TTC ATC ATC GAG
3' DGAT1 (S205A)
~~~~~

Q Y I N P I V R N S K H P L K G D L L Y  
CAG TAC ATC AAC CCT ATT GTG AGA AAC TCT AAG CAC CCT CTC AAG GGT GAT CTT CTC TAC  
~~~~~

3' DGAT1 (S205A)

~~~~~  
A I E R V L K L S V P N L Y V W L C M F  
GCT ATC GAG AGA GTT CTT AAG CTC TCT GTG CCT AAC CTT TAT GTG TGG CTC TGC ATG TTC

3' DGAT1 (S205A)

~~~~~  
Y C F F H L W L N I L A E L L C F G D R
TAC TGT TTC TTC CAC CTC TGG CTT AAC ATC CTT GCT GAG TTG CTT TGC TTC GGA GAT AGA

3' DGAT1 (S205A)

~~~~~  
E F Y K D W W N A K S V G D Y W R M W N  
GAG TTC TAC AAG GAT TGG TGG AAC GCT AAG TCT GTT GGA GAT TAT TGG AGA ATG TGG AAC

3' DGAT1 (S205A)

~~~~~  
M P V H K W M V R H I Y F P C L R S K I
ATG CCT GTG CAT AAG TGG ATG GTG CGT CAC ATC TAC TTC CCT TGC CTC AGA TCT AAG ATC

3' DGAT1 (S205A)

~~~~~  
P K T L A I I I A F L V S A V F H E L C  
CCT AAG ACT CTC GCT ATC ATT ATC GCT TTC CTC GTG TCT GCT GTT TTC CAT GAG TTG TGT

3' DGAT1 (S205A)

~~~~~  
I A V P C R L F K L W A F L G I M F Q V
ATC GCT GTT CCT TGC AGA CTT TTC AAG CTT TGG GCT TTC CTC GGA ATC ATG TTC CAG GTT

3' DGAT1 (S205A)

~~~~~  
P L V F I T N Y L Q E R F G S T V G N M  
CCA CTC GTG TTC ATC ACT AAC TAC CTC CAA GAG AGA TTC GGA TCT ACT GTT GGA AAC ATG

3' DGAT1 (S205A)

~~~~~  
I F W F I F C I F G Q P M C V L L Y Y H
ATT TTC TGG TTC ATT TTC TGC ATC TTC GGA CAG CCT ATG TGC GTT CTC CTC TAC TAC CAC

3' DGAT1 (S205A)

~~~~~  
D L M N R K G S M S \* \*  
GAT CTC ATG AAC AGA AAG GGA TCT ATG TCT TAA TGA AGG ATC CAC CCA GCT TTC TTG TAC

attB2

ocs

~~~~~  
AAA GTG GTG ATG GGT TCG AAA TCG ATA AGC TTG GAT CCT CTA GAG TCC TGC TTT AAT GAG

attB2

ocs

~~~~~  
ATA TGC GAG ACG CCT ATG ATC GCA TGA TAT TTG CTT TCA ATT CTG TTG TGC ACG TTG TAA

ocs

~~~~~  
AAA ACC TGA GCA TGT GTA GCT CAG ATC CTT ACC GCC GGT TTC GGT TCA TTC TAA TGA ATA

ocs

~~~~~  
TAT CAC CCG TTA CTA TCG TAT TTT TAT GAA TAA TAT TCT CCG TTC AAT TTA CTG ATT GTA

ocs

~~~~~  
CCC TAC TAC TTA TAT GTA CAA TAT TAA AAT GAA AAC AAT ATA TTG TGC TGA ATA GGT TTA

ocs

~~~~~  
TAG CGA CAT CTA TGA TAG AGC GCC ACA ATA ACA AAC AAT TGC GTT TTA TTA TTA CAA ATC  
OCS

~~~~~  
CAA TTT TAA AAA AAG CGG CAG AAC CGG TCA AAC CTA AAA GAC TGA TTA CAT AAA TCT TAT
OCS

~~~~~  
TCA AAT TTC AAA AGG CCC CAG GGG CTA GTA TCT ACG ACA CAC CGA GCG GCG AAC TAA TAA  
OCS

~~~~~  
CGT TCA CTG AAG GGA ACT CCG GTT CCC CGC CGG CGC GCA TGG GTG AGA TTC CTT GAA GTT
OCS

~~~~~  
GAG TAT TGG CCG TCC GCT CTA CCG AAA GTT ACG GGC ACC ATT CAA CCC GGT CCA GCA CGG  
OCS

~~~~~  
CGG CCG GGT AAC CGA CTT GCT GCC CCG AGA ATT ATG CAG CAT TTT TTT GGT GTA TGT GGG
OCS

~~~~~  
CCC CAA ATG AAG TGC AGG TCA AAC CTT GAC AGT GAC GAC AAA TCG TTG GGC GGG TCC AGG  
OCS

~~~~~  
GCG AAT TTT GCG ACA ACA TGT CGA GGC TCA GCA GGA CCT GCA GGC ATG CAA GCT AGC TTA
OCS

~~~~~  
CTA GTG ATG CAT ATT CTA TAG TGT CAC CTA AAT CT



## Figure 3

CaMV 35S  
~~~~~  
T CGA CGA ATT AAT TCC AAT CCC ACA AAA ATC
CaMV 35S
~~~~~  
TGA GCT TAA CAG CAC AGT TGC TCC TCT CAG AGC AGA ATC GGG TAT TCA ACA CCC TCA TAT  
CaMV 35S  
~~~~~  
CAA CTA CTA CGT TGT GTA TAA CGG TCC ACA TGC CGG TAT ATA CGA TGA CTG GGG TTG TAC
CaMV 35S
~~~~~  
AAA GGC GGC AAC AAA CGG CGT TCC CGG AGT TGC ACA CAA GAA ATT TGC CAC TAT TAC AGA  
CaMV 35S  
~~~~~  
GGC AAG AGC AGC AGC TGA CGC GTA CAC AAC AAG TCA GCA AAC AGA CAG GTT GAA CTT CAT
CaMV 35S
~~~~~  
CCC CAA AGG AGA AGC TCA ACT CAA GCC CAA GAG CTT TGC TAA GGC CCT AAC AAG CCC ACC  
CaMV 35S  
~~~~~  
AAA GCA AAA AGC CCA CTG GCT CAC GCT AGG AAC CAA AAG GCC CAG CAG TGA TCC AGC CCC
CaMV 35S
~~~~~  
AAA AGA GAT CTC CTT TGC CCC GGA GAT TAC AAT GGA CGA TTT CCT CTA TCT TTA CGA TCT  
CaMV 35S  
~~~~~  
AGG AAG GAA GTT CGA AGG TGA AGG TGA CGA CAC TAT GTT CAC CAC TGA TAA TGA GAA GGT
CaMV 35S
~~~~~  
TAG CCT CTT CAA TTT CAG AAA GAA TGC TGA CCC ACA GAT GGT TAG AGA GGC CTA CGC AGC  
CaMV 35S  
~~~~~  
AGG TCT CAT CAA GAC GAT CTA CCC GAG TAA CAA TCT CCA GGA GAT CAA ATA CCT TCC CAA
CaMV 35S
~~~~~  
GAA GGT TAA AGA TGC AGT CAA AAG ATT CAG GAC TAA TTG CAT CAA GAA CAC AGA GAA AGA  
CaMV 35S  
~~~~~  
CAT ATT TCT CAA GAT CAG AAG TAC TAT TCC AGT ATG GAC GAT TCA AGG CTT GCT TCA TAA
CaMV 35S
~~~~~  
ACC AAG GCA AGT AAT AGA GAT TGG AGT CTC TAA AAA GGT AGT TCC TAC TGA ATC TAA GGC  
CaMV 35S  
~~~~~  
CAT GCA TGG ACT CTA AGA TTC AAA TCG AGG ATC TAA CAG AAC TCG CCG TGA AGA CTG GCG
CaMV 35S
~~~~~  
AAC AGT TCA TAC AGA GTC TTT TAC GAC TCA ATG ACA AGA AGA AAA TCT TCG TCA ACA TGG  
CaMV 35S  
~~~~~  
TGG AGC ACG ACA CTC TGG TCT ACT CCA AAA ATG TCA AAG ATA CAG TCT CAG AAG ACC AAA
CaMV 35S

~~~~~  
GGG CTA TTG AGA CTT TTC AAC AAA GGA TAA TTT CGG GAA ACC TCC TCG GAT TCC ATT GCC  
CaMV 35S

~~~~~  
CAG CTA TCT GTC ACT TCA TCG AAA GGA CAG TAG AAA AGG AAG GTG GCT CCT ACA AAT GCC
CaMV 35S

~~~~~  
ATC ATT GCG ATA AAG GAA AGG CTA TCA TTC AAG ATC TCT CTG CCG ACA GTG GTC CCA AAG  
CaMV 35S

~~~~~  
ATG GAC CCC CAC CCA CGA GGA GCA TCG TGG AAA AAG AAG ACG TTC CAA CCA CGT CTT CAA
CaMV 35S

~~~~~  
AGC AAG TGG ATT GAT GTG ACA TCT CCA CTG ACG TAA GGG ATG ACG CAC AAT CCC ACT ATC  
CaMV 35S

~~~~~  
CTT CGC AAG ACC CTT CCT CTA TAT AAG GAA GTT CAT TTC ATT TGG AGA GGA CAC GCT CGA
attB1

~~~~~  
GGA ATT CGG TAC CCC ATC ACA AGT TTG TAC AAA AAA GCA GGC TGC GGC CGC TTG CTC CCT  
5' Oleo\_1-3

~~~~~  
NcoI

~~~~~  
M A C H Y G Q Q Q Q T R A P H L Q  
TAA AAA AAA CCA TGG CAT GTC ATT ACG GAC AAC AGC AAC AGA CTA GAG CAC CTC ATC TTC  
5' Oleo\_1-3

~~~~~  
UBQ10

~~~~~  
L Q P R A Q R V V K A A T A V T  
AGC TTC AAC CTA GAG CAC AGA GAG TTG TGA AGG CTG CTA CTG CTG TTA CTG TAA ATT TCT  
UBQ10

~~~~~  
GTG TTC CTT ATT CTC TCA AAA TCT TCG ATT TTG TTT TCG TTC GAT CCC AAT TTC GTA TAT
UBQ10

~~~~~  
GTT CTT TGG TTT AGA TTC TGT TAA TCT TAG ATC GAA GAC GAT TTT CTG GGT TTG ATC GTT  
UBQ10

~~~~~  
AGA TAT CAT CTT AAT TCT CGA TTA GGG TTT CAT AGA TAT CAT CCG ATT TGT TCA AAT AAT
UBQ10

~~~~~  
TTG AGT TTT GTC GAA TAA TTA CTC TTC GAT TTG TGA TTT CTA TCT AGA TCT GGT GTT AGT  
3'Oleo\_1-3

~~~~~  
UBQ10

~~~~~  
A G  
TTC TAG TTT GTG CGA TCG AAT TTG TCG ATT AAT CTG AGT TTT TCT GAT TAA CAG GCT CGA  
3' Oleo\_1-3

~~~~~  
G S L L V L S G L T L A G T V I A L T I
GGA TCT CTT CTT GTT CTC TCT GGA CTT ACT CTC GCT GGA ACT GTT ATC GCT CTC ACT ATC
3' Oleo_1-3

~~~~~  
A T P L L V I F S P V L V P A V I T I F  
GCT ACA CCT CTT CTC GTT ATC TTC TCT CCT GTT CTC GTT CCT GCT GTG ATC ACT ATC TTC  
3' Oleo\_1-3  
~~~~~

L L G A G F L A S G G F G V A A L S V L
CTT CTC GGA GCT GGA TTT CTT GCT TCT GGT GGA TTT GGA GTT GCT GCT CTC TCT GTT CTC
3' Oleo_1-3
~~~~~

S W I Y R Y L T G K H P P G A D C L E S  
TCT TGG ATC TAC AGA TAC CTC ACT GGA AAA CAT CCT CCA GGT GCT GAT TGT CTT GAG TCT  
3' Oleo\_1-3  
~~~~~

A K T K L A S C A R E M K D R A E Q F S
GCT AAG ACT AAG CTC GCT TCT TGT GCT AGA GAG ATG AAG GAT AGA GCA GAG CAA TTC TCT
3' Oleo_1-3 OCS terminator
~~~~~

C Q P V A G S Q T S \* \*  
TGT CAG CCT GTT GCT GGA TCT CAG ACT TCT TAA TGA ACA TAT GGT CCT GCT TTA ATG AGA  
OCS terminator  
~~~~~

TAT GCG AGA CGC CTA TGA TCG CAT GAT ATT TGC TTT CAA TTC TGT TGT GCA CGT TGT AAA
OCS terminator
~~~~~

AAA CCT GAG CAT GTG TAG CTC AGA TCC TTA CCG CCG CTT TCG GTT CAT TCT AAT GAA TAT  
OCS terminator  
~~~~~

ATC ACC CGT TAC TAT CGT ATT TTT ATG AAT AAT ATT CTC CGT TCA ATT TAC TGA TTG TAC
OCS terminator
~~~~~

CCT ACT ACT TAT ATG TAC AAT ATT AAA ATG AAA ACA ATA TAT TGT GCT GAA TAG GTT TAT  
OCS terminator  
~~~~~

AGC GAC ATC TAT GAT AGA GCG CCA CAA TAA CAA ACA ATT GCG TTT TAT TAT TAC AAA TCC
OCS terminator
~~~~~

AAT TTT AAA AAA AGC GGC AGA ACC GGT CAA ACC TAA AAG ACT GAT TAC ATA AAT CTT ATT  
OCS terminator  
~~~~~

CAA ATT TCA AAA GGC CCC AGG GGC TAG TAT CTA CGA CAC ACC GAG CGG CGA ACT AAT AAC
OCS terminator
~~~~~

GTT CAC TGA AGG GAA CTC CGG TTC CCC GCC GGC GCG CAT GGG TGA GAT TCC TTG AAG TTG  
OCS terminator  
~~~~~

AGT ATT GGC CGT CCG CTC TAC CGA AAG TTA CGG GCA CCA TTC AAC CCG GTC CAG CAC GGC
OCS terminator
~~~~~

GGC CGG GTA ACC GAC TTG CTG CCC CGA GAA TTA TGC AGC ATT TTT TTG GTG TAT GTG GGC  
OCS terminator  
~~~~~

CCC AAA TGA AGT GCA GGT CAA ACC TTG ACA GTG ACG ACA AAT CGT TGG CCG GGT CCA GGC
OCS terminator
~~~~~

CGA ATT TTG CGA CAA CAT GTC GAG GCT CAG CAG GAC CTG CAG GCA TGC AAG CTA GCT TAC  
~~~~~

OCS terminator

CaMV35S

TAG TGA TGC ATA TTC TAT AGT GTC ACC TAA ATC TTC GAC GAA TTA ATT CCA ATC CCA CAA

CaMV35S

AAA TCT GAG CTT AAC AGC ACA GTT GCT CCT CTC AGA GCA GAA TCG GGT ATT CAA CAC CCT

CaMV35S

CAT ATC AAC TAC TAC GTT GTG TAT AAC GGT CCA CAT GCC GGT ATA TAC GAT GAC TGG GGT

CaMV35S

TGT ACA AAG GCG GCA ACA AAC GGC GTT CCC GGA GTT GCA CAC AAG AAA TTT GCC ACT ATT

CaMV35S

ACA GAG GCA AGA GCA GCA GCT GAC GCG TAC ACA ACA AGT CAG CAA ACA GAC AGG TTG AAC

CaMV35S

TTC ATC CCC AAA GGA GAA GCT CAA CTC AAG CCC AAG AGC TTT GCT AAG GCC CTA ACA AGC

CaMV35S

CCA CCA AAG CAA AAA GCC CAC TGG CTC ACG CTA GGA ACC AAA AGG CCC AGC AGT GAT CCA

CaMV35S

GCC CCA AAA GAG ATC TCC TTT GCC CCG GAG ATT ACA ATG GAC GAT TTC CTC TAT CTT TAC

CaMV35S

GAT CTA GGA AGG AAG TTC GAA GGT GAA GGT GAC GAC ACT ATG TTC ACC ACT GAT AAT GAG

CaMV35S

AAG GTT AGC CTC TTC AAT TTC AGA AAG AAT GCT GAC CCA CAG ATG GTT AGA GAG GCC TAC

CaMV35S

GCA GCA GGT CTC ATC AAG ACG ATC TAC CCG AGT AAC AAT CTC CAG GAG ATC AAA TAC CTT

CaMV35S

CCC AAG AAG GTT AAA GAT GCA GTC AAA AGA TTC AGG ACT AAT TGC ATC AAG AAC ACA GAG

CaMV35S

AAA GAC ATA TTT CTC AAG ATC AGA AGT ACT ATT CCA GTA TGG ACG ATT CAA GGC TTG CTT

CaMV35S

CAT AAA CCA AGG CAA GTA ATA GAG ATT GGA GTC TCT AAA AAG GTA GTT CCT ACT GAA TCT

CaMV35S

AAG GCC ATG CAT GGA GTC TAA GAT TCA AAT CGA GGA TCT AAC AGA ACT CGC CGT GAA GAC

CaMV35S

TGG CGA ACA GTT CAT ACA GAG TCT TTT ACG ACT CAA TGA CAA GAA GAA AAT CTT CGT CAA

CaMV35S

CAT GGT GGA GCA CGA CAC TCT GGT CTA CTC CAA AAA TGT CAA AGA TAC AGT CTC AGA AGA

CaMV35S

CCA AAG GGC TAT TGA GAC TTT TCA ACA AAG GAT AAT TTC GGG AAA CCT CCT CGG ATT CCA
CaMV35S

TTG CCC AGC TAT CTG TCA CTT CAT CGA AAG GAC AGT AGA AAA GGA AGG TGG CTC CTA CAA
CaMV35S

ATG CCA TCA TTG CGA TAA AGG AAA GGC TAT CAT TCA AGA TCT CTC TGC CGA CAG TGG TCC
CaMV35S

CAA AGA TGG ACC CCC ACC CAC GAG GAG CAT CGT GGA AAA AGA AGA CGT TCC AAC CAC GTC
CaMV35S

TTC AAA GCA AGT GGA TTG ATG TGA CAT CTC CAC TGA CGT AAG GGA TGA CGC ACA ATC CCA
CaMV35S

CTA TCC TTC GCA AGA CCC TTC CTC TAT ATA AGG AAG TTC ATT TCA TTT GGA GAG GAC ACG
5' DGAT1 (S205A)

NcoI

M A I L D S A G V T T V
GGA TCC TTG CTC CGT TAA AAA AAA CCA TGG CTA TCC TCG ATT CTG CTG GTG TTA CTA CTG
5' DGAT1 (S205A)

T E N G G E F V D L D R L R R R K S R
TGA CTG AGA ATG GTG GTG GAG AGT TCG TTG ATC TCG ATA GAC TCA GAA GAA GAA AGT CTA
5' DGAT1 (S205A)

UBQ10

S
GAT CTG TAA ATT TCT GTG TTC CTT ATT CTC TCA AAA TCT TCG ATT TTG TTT TCG TTC GAT
UBQ10

CCC AAT TTC GTA TAT GTT CTT TGG TTT AGA TTC TGT TAA TCT TAG ATC GAA GAC GAT TTT
UBQ10

CTG GGT TTG ATC GTT AGA TAT CAT CTT AAT TCT CGA TTA GGG TTT CAT AGA TAT CAT CCG
UBQ10

ATT TGT TCA AAT AAT TTG AGT TTT GTC GAA TAA ITA CTC TTC GAT TTG TGA TTT CTA TCT
UBQ10

AGA TCT GGT GTT AGT TTC TAG TTT GTG CGA TCG AAT TTG TCG ATT AAT CTG AGT TTT TCT
UBQ10

3' DGAT1 (S205A)

D S S N G L L L S G S D N N S P S
GAT TAA CAG GAT TCT TCT AAC GGA CTT CTC CTC TCT GGA TCT GAT AAC AAC TCT CCT TCT
3' DGAT1 (S205A)

D D V G A P A D V R D R I D S V V N D D
GAT GAT GTT GGT GCT CCT GCT GAT GTG AGA GAT AGA ATC GAT TCT GTG GTG AAC GAT GAT
3' DGAT1 (S205A)

~~~~~  
A Q G T A N L A G D N N G G G D N N G G  
GCT CAA GGT ACT GCT AAC CTC GCT GGT GAT AAT AAC GGT GGA GGT GAT AAC AAT GGT GGA  
3' DGAT1 (S205A)  
~~~~~

G R G G G E G R G N A D A T F T Y R P S
GGA AGA GGT GGA GGT GAA GGT AGA GGA AAC GCT GAT GCT ACT TTC ACT TAC AGA CCA TCT
3' DGAT1 (S205A)
~~~~~

V P A H R R A R E S P L S S D A I F K Q  
GTG CCT GCT CAT AGA AGA GCT AGA GAG TCT CCT CTC TCT TCT GAT GCT ATC TTC AAG CAG  
3' DGAT1 (S205A)  
~~~~~

S H A G L F N L C V V V L I A V N S R L
TCT CAC GCT GGA CTT TTC AAC CTC TGT GTG GTG GTT CTT ATC GCT GTG AAC TCT AGA CTC
3' DGAT1 (S205A)
~~~~~

I I E N L M K Y G W L I R T D F W F S S  
ATC ATC GAG AAC CTC ATG AAG TAC GGA TGG CTC ATC AGA ACT GAT TTC TGG TTC TCT TCT  
3' DGAT1 (S205A)  
~~~~~

R S L R D W P L F M C C I S L S I F P L
AGA TCT CTC AGA GAT TGG CCT CTT TTC ATG TGC TGC ATC TCA CTC TCA ATC TTC CCT CTC
3' DGAT1 (S205A)
~~~~~

A A F T V E K L V L Q K Y I A E P V V I  
GCT GCT TTT ACT GTT GAG AAG CTC GTG CTC CAG AAG TAT ATC GCT GAA CCT GTG GTG ATC  
3' DGAT1 (S205A)  
~~~~~

F L H I I I T M T E V L Y P V Y V T L R
TTC CTC CAC ATC ATC ATC ACT ATG ACT GAG GTT CTC TAC CCT GTT TAC GTG ACT CTC AGA
3' DGAT1 (S205A)
~~~~~

C D S A F L S G V T L M L L T C I V W L  
TGC GAT TCT GCT TTC CTC TCT GGT GTT ACT CTT ATG CTC CTC ACT TGC ATT GTG TGG CTT  
3' DGAT1 (S205A)  
~~~~~

K L V S Y A H T S Y D I R S L A N A A D
AAG CTC GTG TCT TAC GCT CAC ACT TCT TAC GAT ATC AGA TCT CTC GCT AAC GCT GCT GAT
3' DGAT1 (S205A)
~~~~~

K A N P E V S Y Y V S L K S L A Y F M V  
AAG GCT AAC CCT GAA GTG TCT TAC TAC GTG TCT CTC AAG TCT CTC GCT TAC TTC ATG GTT  
3' DGAT1 (S205A)  
~~~~~

A P T L C Y Q P S Y P R S A C I R K G W
GCT CCT ACA CTT TGT TAC CAG CCA TCT TAC CCT AGA TCT GCT TGC ATT AGA AAG GGA TGG
3' DGAT1 (S205A)
~~~~~

V A R Q F A K L V I F T G F M G F I I E  
GTG GCA AGA CAA TTC GCT AAG TTG GTG ATC TTC ACT GGA TTC ATG GGA TTC ATC ATC GAG  
3' DGAT1 (S205A)  
~~~~~

Q Y I N P I V R N S K H P L K G D L L Y
CAG TAC ATC AAC CCT ATT GTG AGA AAC TCT AAG CAC CCT CTC AAG GGT GAT CTT CTC TAC
~~~~~

3' DGAT1 (S205A)

~~~~~  
A I E R V L K L S V P N L Y V W L C M F
GCT ATC GAG AGA GTT CTT AAG CTC TCT GTG CCT AAC CTT TAT GTG TGG CTC TGC ATG TTC

3' DGAT1 (S205A)

~~~~~  
Y C F F H L W L N I L A E L L C F G D R  
TAC TGT TTC TTC CAC CTC TGG CTT AAC ATC CTT GCT GAG TTG CTT TGC TTC GGA GAT AGA

3' DGAT1 (S205A)

~~~~~  
E F Y K D W W N A K S V G D Y W R M W N
GAG TTC TAC AAG GAT TGG TGG AAC GCT AAG TCT GTT GGA GAT TAT TGG AGA ATG TGG AAC

3' DGAT1 (S205A)

~~~~~  
M P V H K W M V R H I Y F P C L R S K I  
ATG CCT GTG CAT AAG TGG ATG GTG CGT CAC ATC TAC TTC CCT TGC CTC AGA TCT AAG ATC

3' DGAT1 (S205A)

~~~~~  
P K T L A I I I A F L V S A V F H E L C
CCT AAG ACT CTC GCT ATC ATT ATC GCT TTC CTC GTG TCT GCT GTT TTC CAT GAG TTG TGT

3' DGAT1 (S205A)

~~~~~  
I A V P C R L F K L W A F L G I M F Q V  
ATC GCT GTT CCT TGC AGA CTT TTC AAG CTT TGG GCT TTC CTC GGA ATC ATG TTC CAG GTT

3' DGAT1 (S205A)

~~~~~  
P L V F I T N Y L Q E R F G S T V G N M
CCA CTC GTG TTC ATC ACT AAC TAC CTC CAA GAG AGA TTC GGA TCT ACT GTT GGA AAC ATG

3' DGAT1 (S205A)

~~~~~  
I F W F I F C I F G Q P M C V L L Y Y H  
ATT TTC TGG TTC ATT TTC TGC ATC TTC GGA CAG CCT ATG TGC GTT CTC CTC TAC TAC CAC

3' DGAT1 (S205A)

~~~~~  
D L M N R K G S M S * *
GAT CTC ATG AAC AGA AAG GGA TCT ATG TCT TAA TGA AGG ATC CAC CCA GCT TTC TTG TAC

attB2

ocs

~~~~~  
AAA GTG GTG ATG GGT TCG AAA TCG ATA AGC TTG GAT CCT CTA GAG TCC TGC TTT AAT GAG

attB2

ocs

~~~~~  
ATA TGC GAG ACG CCT ATG ATC GCA TGA TAT TTG CTT TCA ATT CTG TTG TGC ACG TTG TAA

ocs

~~~~~  
AAA ACC TGA GCA TGT GTA GCT CAG ATC CTT ACC GCC GGT TTC GGT TCA TTC TAA TGA ATA

ocs

~~~~~  
TAT CAC CCG TTA CTA TCG TAT TTT TAT GAA TAA TAT TCT CCG TTC AAT TTA CTG ATT GTA

ocs

~~~~~  
CCC TAC TAC TTA TAT GTA CAA TAT TAA AAT GAA AAC AAT ATA TTG TGC TGA ATA GGT TTA

ocs

~~~~~  
TAG CGA CAT CTA TGA TAG AGC GCC ACA ATA ACA AAC AAT TGC GTT TTA TTA TTA CAA ATC

OCS

~~~~~  
CAA TTT TAA AAA AAG CGG CAG AAC CGG TCA AAC CTA AAA GAC TGA TTA CAT AAA TCT TAT

OCS

~~~~~  
TCA AAT TTC AAA AGG CCC CAG GGG CTA GTA TCT ACG ACA CAC CGA GCG GCG AAC TAA TAA

OCS

~~~~~  
CGT TCA CTG AAG GGA ACT CCG GTT CCC CGC CGG CGC GCA TGG GTG AGA TTC CTT GAA GTT

OCS

~~~~~  
GAG TAT TGG CCG TCC GCT CTA CCG AAA GTT ACG GGC ACC ATT CAA CCC GGT CCA GCA CGG

OCS

~~~~~  
CGG CCG GGT AAC CGA CTT GCT GCC CCG AGA ATT ATG CAG CAT TTT TTT GGT GTA TGT GGG

OCS

~~~~~  
CCC CAA ATG AAG TGC AGG TCA AAC CTT GAC AGT GAC GAC AAA TCG TTG GGC GGG TCC AGG

OCS

~~~~~  
GCG AAT TTT GCG ACA ACA TGT CGA GGC TCA GCA GGA CCT GCA GGC ATG CAA GCT AGC TTA

OCS

~~~~~  
CTA GTG ATG CAT ATT CTA TAG TGT CAC CTA AAT CT

Figure 4

CaMV 35S
~~~~~  
TCG ACG AAT TAA TTC CAA TCC CAC AAA AAT  
~~~~~  
CaMV 35S
~~~~~  
CTG AGC TTA ACA GCA CAG TTG CTC CTC TCA GAG CAG AAT CGG GTA TTC AAC ACC CTC ATA  
~~~~~  
CaMV 35S
~~~~~  
TCA ACT ACT ACG TTG TGT ATA ACG GTC CAC ATG CCG GTA TAT ACG ATG ACT GGG GTT GTA  
~~~~~  
CaMV 35S
~~~~~  
CAA AGG CGG CAA CAA ACC GCG TTC CCG GAG TTG CAC ACA AGA AAT TTG CCA CTA TTA CAG  
~~~~~  
CaMV 35S
~~~~~  
AGG CAA GAG CAG CAG CTG ACG CGT ACA CAA CAA GTC AGC AAA CAG ACA GGT TGA ACT TCA  
~~~~~  
CaMV 35S
~~~~~  
TCC CCA AAG GAG AAG CTC AAC TCA AGC CCA AGA GCT TTG CTA AGG CCC TAA CAA GCC CAC  
~~~~~  
CaMV 35S
~~~~~  
CAA AGC AAA AAG CCC ACT GGC TCA CGC TAG GAA CCA AAA GGC CCA GCA GTG ATC CAG CCC  
~~~~~  
CaMV 35S
~~~~~  
CAA AAG AGA TCT CCT TTG CCC CGG AGA TTA CAA TGG ACG ATT TCC TCT ATC TTT ACG ATC  
~~~~~  
CaMV 35S
~~~~~  
TAG GAA GGA AGT TCG AAG GTG AAG GTG ACG ACA CTA TGT TCA CCA CTG ATA ATG AGA AGG  
~~~~~  
CaMV 35S
~~~~~  
TTA GCC TCT TCA ATT TCA GAA AGA ATG CTG ACC CAC AGA TGG TTA GAG AGG CCT ACG CAG  
~~~~~  
CaMV 35S
~~~~~  
CAG GTC TCA TCA AGA CGA TCT ACC CGA GTA ACA ATC TCC AGG AGA TCA AAT ACC TTC CCA  
~~~~~  
CaMV 35S
~~~~~  
AGA AGG TTA AAG ATG CAG TCA AAA GAT TCA GGA CTA ATT GCA TCA AGA ACA CAG AGA AAG  
~~~~~  
CaMV 35S
~~~~~  
ACA TAT TTC TCA AGA TCA GAA GTA CTA TTC CAG TAT GGA CGA TTC AAG GCT TGC TTC ATA  
~~~~~  
CaMV 35S
~~~~~  
AAC CAA GGC AAG TAA TAG AGA TTG GAG TCT CTA AAA AGG TAG TTC CTA CTG AAT CTA AGG  
~~~~~  
CaMV 35S
~~~~~  
CCA TGC ATG GAG TCT AAG ATT CAA ATC GAG GAT CTA ACA GAA CTC GCC GTG AAG ACT GGC  
~~~~~  
CaMV 35S
~~~~~  
GAA CAG TTC ATA CAG AGT CTT TTA CGA CTC AAT GAC AAG AAG AAA ATC TTC GTC AAC ATG  
~~~~~  
CaMV 35S
~~~~~  
GTG GAG CAC GAC ACT CTG GTC TAC TCC AAA AAT GTC AAA GAT ACA GTC TCA GAA GAC CAA  
~~~~~  
CaMV 35S

~~~~~  
AGG GCT ATT GAG ACT TTT CAA CAA AGG ATA ATT TCG GGA AAC CTC CTC GGA TTC CAT TGC  
CaMV 35S

~~~~~  
CCA GCT ATC TGT CAC TTC ATC GAA AGG ACA GTA GAA AAG GAA GGT GGC TCC TAC AAA TGC
CaMV 35S

~~~~~  
CAT CAT TGC GAT AAA GGA AAG GCT ATC ATT CAA GAT CTC TCT GCC GAC AGT GGT CCC AAA  
CaMV 35S

~~~~~  
GAT GGA CCC CCA CCC ACG AGG AGC ATC GTG GAA AAA GAA GAC GTT CCA ACC ACG TCT TCA
CaMV 35S

~~~~~  
AAG CAA GTG GAT TGA TGT GAC ATC TCC ACT GAC GTA AGG GAT GAC GCA CAA TCC CAC TAT  
CaMV 35S

~~~~~  
CCT TCG CAA GAC CCT TCC TCT ATA TAA GGA AGT TCA TTT CAT TTG GAG AGG ACA CGC TCG
attB1

~~~~~  
AGG AAT TCG GTA CCC CAT CAC AAG TTT GTA CAA AAA AGC AGG CTG CGG CCG CTT GCT CCC  
5' Oleo\_3-1

~~~~~  
NcoI

~~~~~  
M A C H Y G Q Q Q Q T C A P H L  
TTA AAA AAA ACC ATG GCA TGT CAT TAC GGA CAA CAG CAA CAG ACT TGT GCA CCT CAT CTT  
5' Oleo\_3-1

~~~~~  
UBQ10

~~~~~  
Q L Q P R A C R V V K A A T A V T  
CAG CTT CAA CCT AGA GCA TGT AGA GTT GTG AAG GCT GCT ACT GCT GTT ACT GTA AAT TTC  
UBQ10

~~~~~  
TGT GTT CCT TAT TCT CTC AAA ATC TTC GAT TTT GTT TTC GTT CGA TCC CAA TTT CGT ATA
UBQ10

~~~~~  
TGT TCT TTG GTT TAG ATT CTG TTA ATC TTA GAT CGA AGA CGA TTT TCT GGG TTT GAT CGT  
UBQ10

~~~~~  
TAG ATA TCA TCT TAA TTC TCG ATT AGG GTT TCA TAG ATA TCA TCC GAT TTG TTC AAA TAA
UBQ10

~~~~~  
TTT GAG TTT TGT CGA ATA ATT ACT CTT CGA TTT GTG ATT TCT ATC TAG ATC TGG TGT TAG  
3'Oleo\_3-1

~~~~~  
UBQ10

~~~~~  
A G  
TTT CTA GTT TGT GCG ATC GAA TTT GTC GAT TAA TCT GAG TTT TTC TGA TTA ACA GGC TGG  
3' Oleo\_3-1

~~~~~  
G S L L V L S G L T L A G T V I A L T I
AGG ATC TCT TCT TGT TCT CTC TGG ACT TAC TCT CGC TGG AAC TGT TAT CGC TCT CAC TAT
3' Oleo_3-1

~~~~~  
· A T P L L V I F S P V L V P A V I T I F  
CGC TAC ACC TCT TCT CGT TAT CTT CTC TCC TGT TCT CGT TCC TGC TGT GAT CAC TAT CTT  
3' Oleo\_3-1  
~~~~~

· L L G A G F L A S G G F G V A A L S V L
CCT TCT CGG AGC TGG ATT TCT TGC TTC TGG TGG ATT TGG AGT TGC TGC TCT CTC TGT TCT
3' Oleo_3-1
~~~~~

· S W I Y R Y L T G K H P P G A D Q L E S  
CTC TTG GAT CTA CAG ATA CCT CAC TGG AAA ACA TCC TCC AGG TGC TGA TCA ACT TGA GTC  
3' Oleo\_3-1  
~~~~~

· A K T K L A S K A R E M K D R A E Q F S
TGC TAA GAC TAA GCT CGC TTC TAA GGC TAG AGA GAT GAA GGA TAG AGC AGA GCA ATT CTC
3' Oleo_3-1 OCS terminator
~~~~~

· C Q P V A G S Q T S \* \*  
TTG TCA GCC TGT TGC TGG ATC TCA GAC TTC TTA ATG AAC ATA TGG TCC TGC TTT AAT GAG  
OCS terminator  
~~~~~

ATA TGC GAG ACG CCT ATG ATC GCA TGA TAT TTG CTT TCA ATT CTG TTG TGC ACG TTG TAA
OCS terminator
~~~~~

AAA ACC TGA GCA TGT GTA GCT CAG ATC CTT ACC GCC GGT TTC GGT TCA TTC TAA TGA ATA  
OCS terminator  
~~~~~

TAT CAC CCG TTA CTA TCG TAT TTT TAT GAA TAA TAT TCT CCG TTC AAT TTA CTG ATT GTA
OCS terminator
~~~~~

CCC TAC TAC TTA TAT GTA CAA TAT TAA AAT GAA AAC AAT ATA TTG TGC TGA ATA GGT TTA  
OCS terminator  
~~~~~

TAG CGA CAT CTA TGA TAG AGC GCC ACA ATA ACA AAC AAT TGC GTT TTA TTA CAA ATC
OCS terminator
~~~~~

CAA TTT TAA AAA AAG CGG CAG AAC CGG TCA AAC CTA AAA GAC TGA TTA CAT AAA TCT TAT  
OCS terminator  
~~~~~

TCA AAT TTC AAA AGG CCC CAG GGG CTA GTA TCT ACG ACA CAC CGA GCG GCG AAC TAA TAA
OCS terminator
~~~~~

CGT TCA CTG AAG GGA ACT CCG GTT CCC CGC CGC GCA TGG GTG AGA TTC CTT GAA GTT  
OCS terminator  
~~~~~

GAG TAT TGG CCG TCC GCT CTA CCG AAA GTT ACG GGC ACC ATT CAA CCC GGT CCA GCA CGG
OCS terminator
~~~~~

CGG CCG GGT AAC CGA CTT GCT GCC CCG AGA ATT ATG CAG CAT TTT TTT GGT GTA TGT GGG  
OCS terminator  
~~~~~

CCC CAA ATG AAG TGC AGG TCA AAC CTT GAC AGT GAC GAC AAA TCG TTG GGC GGG TCC AGG
OCS terminator
~~~~~

CGG AAT TTT GCG ACA ACA TGT CGA GGC TCA GCA GGA CCT GCA GGC ATG CAA GCT AGC TTA  
~~~~~

OCS terminator

CaMV35S

CTA GTG ATG CAT ATT CTA TAG TGT CAC CTA AAT CTT CGA CGA ATT AAT TCC AAT CCC ACA

CaMV35S

AAA ATC TGA GCT TAA CAG CAC AGT TGC TCC TCT CAG AGC AGA ATC GGG TAT TCA ACA CCC

CaMV35S

TCA TAT CAA CTA CTA CGT TGT GTA TAA CGG TCC ACA TGC CGG TAT ATA CGA TGA CTG GGG

CaMV35S

TTG TAC AAA GGC GGC AAC AAA CGG CGT TCC CGG AGT TGC ACA CAA GAA ATT TGC CAC TAT

CaMV35S

TAC AGA GGC AAG AGC AGC AGC TGA CGC GTA CAC AAC AAG TCA GCA AAC AGA CAG GTT GAA

CaMV35S

CTT CAT CCC CAA AGG AGA AGC TCA ACT CAA GCC CAA GAG CTT TGC TAA GGC CCT AAC AAG

CaMV35S

CCC ACC AAA GCA AAA AGC CCA CTG GCT CAC GCT AGG AAC CAA AAG GCC CAG CAG TGA TCC

CaMV35S

AGC CCC AAA AGA GAT CTC CTT TGC CCC GGA GAT TAC AAT GGA CGA TTT CCT CTA TCT TTA

CaMV35S

CGA TCT AGG AAG GAA GTT CGA AGG TGA AGG TGA CGA CAC TAT GTT CAC CAC TGA TAA TGA

CaMV35S

GAA GGT TAG CCT CTT CAA TTT CAG AAA GAA TGC TGA CCC ACA GAT GGT TAG AGA GGC CTA

CaMV35S

CGC AGC AGG TCT CAT CAA GAC GAT CTA CCC GAG TAA CAA TCT CCA GGA GAT CAA ATA CCT

CaMV35S

TCC CAA GAA GGT TAA AGA TGC AGT CAA AAG ATT CAG GAC TAA TTG CAT CAA GAA CAC AGA

CaMV35S

GAA AGA CAT ATT TCT CAA GAT CAG AAG TAC TAT TCC AGT ATG GAC GAT TCA AGG CTT GCT

CaMV35S

TCA TAA ACC AAG GCA AGT AAT AGA GAT TGG AGT CTC TAA AAA GGT AGT TCC TAC TGA ATC

CaMV35S

TAA GGC CAT GCA TGG AGT CTA AGA TTC AAA TCG AGG ATC TAA CAG AAC TCG CCG TGA AGA

CaMV35S

CTG GCG AAC AGT TCA TAC AGA GTC TTT TAC GAC TCA ATG ACA AGA AGA AAA TCT TCG TCA

CaMV35S

ACA TGG TGG AGC ACG ACA CTC TGG TCT ACT CCA AAA ATG TCA AAG ATA CAG TCT CAG AAG

CaMV35S

ACC AAA GGG CTA TTG AGA CTT TTC AAC AAA GGA TAA TTT CGG GAA ACC TCC TCG GAT TCC
CaMV35S

ATT GCC CAG CTA TCT GTC ACT TCA TCG AAA GGA CAG TAG AAA AGG AAG GTG GCT CCT ACA
CaMV35S

AAT GCC ATC ATT GCG ATA AAG GAA AGG CTA TCA TTC AAG ATC TCT CTG CCG ACA GTG GTC
CaMV35S

CCA AAG ATG GAC CCC CAC CCA CGA GGA GCA TCG TGG AAA AAG AAG ACG TTC CAA CCA CGT
CaMV35S

CTT CAA AGC AAG TGG ATT GAT GTG ACA TCT CCA CTG ACG TAA GGG ATG ACG CAC AAT CCC
CaMV35S

ACT ATC CTT CGC AAG ACC CTT CCT CTA TAT AAG GAA GTT CAT TTC ATT TGG AGA GGA CAC
CaMV35S 5' DGAT1 (S205A)

NcoI

GGG ATC CTT GCT CCG TTA AAA AAA ACC ATG GCT ATC CTC GAT TCT GCT GGT GTT ACT ACT
5' DGAT1 (S205A)

V T E N G G G E F V D L D R L R R R K S
GTG ACT GAG AAT GGT GGT GGA GAG TTC GTT GAT CTC GAT AGA CTC AGA AGA AGA AAG TCT
5' DGAT1 (S205A)

UBQ10

R S
AGA TCT GTA AAT TTC TGT GTT CCT TAT TCT CTC AAA ATC TTC GAT TTT GTT TTC GTT CGA
UBQ10

TCC CAA TTT CGT ATA TGT TCT TTG GTT TAG ATT CTG TTA ATC TTA GAT CGA AGA CGA TTT
UBQ10

TCT GGG TTT GAT CGT TAG ATA TCA TCT TAA TTC TCG ATT AGG GTT TCA TAG ATA TCA TCC
UBQ10

GAT TTG TTC AAA TAA TTT GAG TTT TGT CGA ATA ATT ACT CTT CGA TTT GTG ATT TCT ATC
UBQ10

TAG ATC TGG TGT TAG TTT CTA GTT TGT GCG ATC GAA TTT GTC GAT TAA TCT GAG TTT TTC
UBQ10

3' DGAT1 (S205A)

D S S N G L L L S G S D N N S P S
TGA TTA ACA GGA TTC TTC TAA CGG ACT TCT CCT CTC TGG ATC TGA TAA CAA CTC TCC TTC
3' DGAT1 (S205A)

D D V G A P A D V R D R I D S V V N D D
TGA TGA TGT TGG TGC TCC TGC TGA TGT GAG AGA TAG AAT CGA TTC TGT GGT GAA CGA TGA
3' DGAT1 (S205A)

~~~~~  
· A Q G T A N L A G D N N G G G D N N G G  
TGC TCA AGG TAC TGC TAA CCT CGC TGG TGA TAA TAA CGG TGG AGG TGA TAA CAA TGG TGG  
3' DGAT1 (S205A)  
~~~~~

· G R G G G E G R G N A D A T F T Y R P S
AGG AAG AGG TGG AGG TGA AGG TAG AGG AAA CGC TGA TGC TAC TTT CAC TTA CAG ACC ATC
3' DGAT1 (S205A)
~~~~~

· V P A H R R A R E S P L S S D A I F K Q  
TGT GCC TGC TCA TAG AAG AGC TAG AGA GTC TCC TCT CTC TTC TGA TGC TAT CTT CAA GCA  
3' DGAT1 (S205A)  
~~~~~

· S H A G L F N L C V V V L I A V N S R L
GTC TCA CGC TGG ACT TTT CAA CCT CTG TGT GGT GGT TCT TAT CGC TGT GAA CTC TAG ACT
3' DGAT1 (S205A)
~~~~~

· I I E N L M K Y G W L I R T D F W F S S  
CAT CAT CGA GAA CCT CAT GAA GTA CGG ATG GCT CAT CAG AAC TGA TTT CTG GTT CTC TTC  
3' DGAT1 (S205A)  
~~~~~

· R S L R D W P L F M C C I S L S I F P L
TAG ATC TCT CAG AGA TTG GCC TCT TTT CAT GTG CTG CAT CTC ACT CTC AAT CTT CCC TCT
3' DGAT1 (S205A)
~~~~~

· A A F T V E K L V L Q K Y I A E P V V I  
CGC TGC TTT TAC TGT TGA GAA GCT CGT GCT CCA GAA GTA TAT CGC TGA ACC TGT GGT GAT  
3' DGAT1 (S205A)  
~~~~~

· F L H I I I T M T E V L Y P V Y V T L R
CTT CCT CCA CAT CAT CAT CAC TAT GAC TGA GGT TCT CTA CCC TGT TTA CGT GAC TCT CAG
3' DGAT1 (S205A)
~~~~~

· C D S A F L S G V T L M L L T C I V W L  
ATG CGA TTC TGC TTT CCT CTC TGG TGT TAC TCT TAT GCT CCT CAC TTG CAT TGT GTG GCT  
3' DGAT1 (S205A)  
~~~~~

· K L V S Y A H T S Y D I R S L A N A A D
TAA GCT CGT GTC TTA CGC TCA CAC TTC TTA CGA TAT CAG ATC TCT CGC TAA CGC TGC TGA
3' DGAT1 (S205A)
~~~~~

· K A N P E V S Y Y V S L K S L A Y F M V  
TAA GGC TAA CCC TGA AGT GTC TTA CTA CGT GTC TCT CAA GTC TCT CGC TTA CTT CAT GGT  
3' DGAT1 (S205A)  
~~~~~

· A P T L C Y Q P S Y P R S A C I R K G W
TGC TCC TAC ACT TTG TTA CCA GCC ATC TTA CCC TAG ATC TGC TTG CAT TAG AAA GGG ATG
3' DGAT1 (S205A)
~~~~~

· V A R Q F A K L V I F T G F M G F I I E  
GGT GGC AAG ACA ATT CGC TAA GTT GGT GAT CTT CAC TGG ATT CAT GGG ATT CAT CAT CGA  
3' DGAT1 (S205A)  
~~~~~

· Q Y I N P I V R N S K H P L K G D L L Y
GCA GTA CAT CAA CCC TAT TGT GAG AAA CTC TAA GCA CCC TCT CAA GGG TGA TCT TCT CTA
~~~~~

3' DGAT1 (S205A)

~~~~~  
· A I E R V L K L S V P N L Y V W L C M F
CGC TAT CGA GAG AGT TCT TAA GCT CTC TGT GCC TAA CCT TTA TGT GTG GCT CTG CAT GTT

3' DGAT1 (S205A)

~~~~~  
· Y C F F H L W L N I L A E L L C F G D R  
CTA CTG TTT CTT CCA CCT CTG GCT TAA CAT CCT TGC TGA GTT GCT TTG CTT CGG AGA TAG

3' DGAT1 (S205A)

~~~~~  
· E F Y K D W W N A K S V G D Y W R M W N
AGA GTT CTA CAA GGA TTG GTG GAA CGC TAA GTC TGT TGG AGA TTA TTG GAG AAT GTG GAA

3' DGAT1 (S205A)

~~~~~  
· M P V H K W M V R H I Y F P C L R S K I  
CAT GCC TGT GCA TAA GTG GAT GGT GCG TCA CAT CTA CTT CCC TTG CCT CAG ATC TAA GAT

3' DGAT1 (S205A)

~~~~~  
· P K T L A I I I A F L V S A V F H E L C
CCC TAA GAC TCT CGC TAT CAT TAT CGC TTT CCT CGT GTC TGC TGT TTT CCA TGA GTT GTG

3' DGAT1 (S205A)

~~~~~  
· I A V P C R L F K L W A F L G I M F Q V  
TAT CGC TGT TCC TTG CAG ACT TTT CAA GCT TTG GGC TTT CCT CGG AAT CAT GTT CCA GGT

3' DGAT1 (S205A)

~~~~~  
· P L V F I T N Y L Q E R F G S T V G N M
TCC ACT CGT GTT CAT CAC TAA CTA CCT CCA AGA GAG ATT CGG ATC TAC TGT TGG AAA CAT

3' DGAT1 (S205A)

~~~~~  
· I F W F I F C I F G Q P M C V L L Y Y H  
GAT TTT CTG GTT CAT TTT CTG CAT CTT CGG ACA GCC TAT GTG CGT TCT CCT CTA CTA CCA

3' DGAT1 (S205A)

~~~~~  
· D L M N R K G S M S * *
CGA TCT CAT GAA CAG AAA GGG ATC TAT GTC TTA ATG AAG GAT CCA CCC AGC TTT CTT GTA

attB2

ocs

~~~~~  
CAA AGT GGT GAT GGG TTC GAA ATC GAT AAG CTT GGA TCC TCT AGA GTC CTG CTT TAA TGA

attB2

ocs

~~~~~  
GAT ATG CGA GAC GCC TAT GAT CGC ATG ATA TTT GCT TTC AAT TCT GTT GTG CAC GTT GTA

ocs

~~~~~  
AAA AAC CTG AGC ATG TGT AGC TCA GAT CCT TAC CGC CGG TTT CGG TTC ATT CTA ATG AAT

ocs

~~~~~  
ATA TCA CCC GTT ACT ATC GTA TTT TTA TGA ATA ATA TTC TCC GTT CAA TTT ACT GAT TGT

ocs

~~~~~  
ACC CTA CTA CTT ATA TGT ACA ATA TTA AAA TGA AAA CAA TAT ATT GTG CTG AAT AGG TTT

ocs

~~~~~  
ATA GCG ACA TCT ATG ATA GAG CGC CAC AAT AAC AAA CAA TTG CGT TTT ATT ATT ACA AAT

OCS

~~~~~  
CCA ATT TTA AAA AAA GCG GCA GAA CCG GTC AAA CCT AAA AGA CTG ATT ACA TAA ATC TTA

OCS

~~~~~  
TTC AAA TTT CAA AAG GCC CCA GGG GCT AGT ATC TAC GAC ACA CCG AGC GGC GAA CTA ATA

OCS

~~~~~  
ACG TTC ACT GAA GGG AAC TCC GGT TCC CCG CCG GCG CGC ATG GGT GAG ATT CCT TGA AGT

OCS

~~~~~  
TGA GTA TTG GCC GTC CGC TCT ACC GAA AGT TAC GGG CAC CAT TCA ACC CGG TCC AGC ACG

OCS

~~~~~  
GCG GCC GGG TAA CCG ACT TGC TGC CCC GAG AAT TAT GCA GCA TTT TTT TGG TGT ATG TGG

OCS

~~~~~  
GCC CCA AAT GAA GTG CAG GTC AAA CCT TGA CAG TGA CGA CAA ATC GTT GGG CGG GTC CAG

OCS

~~~~~  
GGC GAA TTT TGC GAC AAC ATG TCG AGG CTC AGC AGG ACC TGC AGG CAT GCA AGC TAG CTT

OCS

~~~~~  
ACT AGT GAT GCA TAT TCT ATA GTG TCA CCT AAA TCT

Figure 5

CaMV 35S
~~~~~  
T CGA CGA ATT AAT TCC AAT CCC ACA AAA ATC  
~~~~~  
CaMV 35S
~~~~~  
TGA GCT TAA CAG CAC AGT TGC TCC TCT CAG AGC AGA ATC GGG TAT TCA ACA CCC TCA TAT  
~~~~~  
CaMV 35S
~~~~~  
CAA CTA CTA CGT TGT GTA TAA CGG TCC ACA TGC CGG TAT ATA CGA TGA CTG GGG TTG TAC  
~~~~~  
CaMV 35S
~~~~~  
AAA GGC GGC AAC AAA CGG CGT TCC CGG AGT TGC ACA CAA GAA ATT TGC CAC TAT TAC AGA  
~~~~~  
CaMV 35S
~~~~~  
GGC AAG AGC AGC AGC TGA CGC GTA CAC AAC AAG TCA GCA AAC AGA CAG GTT GAA CTT CAT  
~~~~~  
CaMV 35S
~~~~~  
CCC CAA AGG AGA AGC TCA ACT CAA GCC CAA GAG CTT TGC TAA GGC CCT AAC AAG CCC ACC  
~~~~~  
CaMV 35S
~~~~~  
AAA GCA AAA AGC CCA CTG GCT CAC GCT AGG AAC CAA AAG GCC CAG CAG TGA TCC AGC CCC  
~~~~~  
CaMV 35S
~~~~~  
AAA AGA GAT CTC CTT TGC CCC GGA GAT TAC AAT GGA CGA TTT CCT CTA TCT TTA CGA TCT  
~~~~~  
CaMV 35S
~~~~~  
AGG AAG GAA GTT CGA AGG TGA AGG TGA CGA CAC TAT GTT CAC CAC TGA TAA TGA GAA GGT  
~~~~~  
CaMV 35S
~~~~~  
TAG CCT CTT CAA TTT CAG AAA GAA TGC TGA CCC ACA GAT GGT TAG AGA GGC CTA CGC AGC  
~~~~~  
CaMV 35S
~~~~~  
AGG TCT CAT CAA GAC GAT CTA CCC GAG TAA CAA TCT CCA GGA GAT CAA ATA CCT TCC CAA  
~~~~~  
CaMV 35S
~~~~~  
GAA GGT TAA AGA TGC AGT CAA AAG ATT CAG GAC TAA TTG CAT CAA GAA CAC AGA GAA AGA  
~~~~~  
CaMV 35S
~~~~~  
CAT ATT TCT CAA GAT CAG AAG TAC TAT TCC AGT ATG GAC GAT TCA AGG CTT GCT TCA TAA  
~~~~~  
CaMV 35S
~~~~~  
ACC AAG GCA AGT AAT AGA GAT TGG AGT CTC TAA AAA GGT AGT TCC TAC TGA ATC TAA GGC  
~~~~~  
CaMV 35S
~~~~~  
CAT GCA TGG AGT CTA AGA TTC AAA TCG AGG ATC TAA CAG AAC TCG CCG TGA AGA CTG GCG  
~~~~~  
CaMV 35S
~~~~~  
AAC AGT TCA TAC AGA GTC TTT TAC GAC TCA ATG ACA AGA AGA AAA TCT TCG TCA ACA TGG  
~~~~~  
CaMV 35S
~~~~~  
TGG AGC ACG ACA CTC TGG TCT ACT CCA AAA ATG TCA AAG ATA CAG TCT CAG AAG ACC AAA  
~~~~~  
CaMV 35S

GGG CTA TTG AGA CTT TTC AAC AAA GGA TAA TTT CGG GAA ACC TCC TCG GAT TCC ATT GCC
CaMV 35S

CAG CTA TCT GTC ACT TCA TCG AAA GGA CAG TAG AAA AGG AAG GTG GCT CCT ACA AAT GCC
CaMV 35S

ATC ATT GCG ATA AAG GAA AGG CTA TCA TTC AAG ATC TCT CTG CCG ACA GTG GTC CCA AAG
CaMV 35S

ATG GAC CCC CAC CCA CGA GGA GCA TCG TGG AAA AAG AAG ACG TTC CAA CCA CGT CTT CAA
CaMV 35S

AGC AAG TGG ATT GAT GTG ACA TCT CCA CTG ACG TAA GGG ATG ACG CAC AAT CCC ACT ATC
CaMV 35S

CTT CGC AAG ACC CTT CCT CTA TAT AAG GAA GTT CAT TTC ATT TGG AGA GGA CAC GCT CGA
attB1

GGA ATT CGG TAC CCC ATC ACA AGT TTG TAC AAA AAA GCA GGC TGC GGC CGC TTG CTC CCT
5' Oleo_3-3

NcoI

M A C H Y G Q Q Q Q T C A P H L Q
TAA AAA AAA CCA TGG CAT GTC ATT ACG GAC AAC AGC AAC AGA CTT GTG CAC CTC ATC TTC
5' Oleo_3-3

UBQ10

L Q P R A C R V V K A A T A V T
AGC TTC AAC CTA GAG CAT GTA GAG TTG TGA AGG CTG CTA CTG CTG TTA CTG TAA ATT TCT
UBQ10

GTG TTC CTT ATT CTC TCA AAA TCT TCG ATT TTG TTT TCG TTC GAT CCC AAT TTC GTA TAT
UBQ10

GTT CTT TGG TTT AGA TTC TGT TAA TCT TAG ATC GAA GAC GAT TTT CTG GGT TTG ATC GTT
UBQ10

AGA TAT CAT CTT AAT TCT CGA TTA GGG TTT CAT AGA TAT CAT CCG ATT TGT TCA AAT AAT
UBQ10

TTG AGT TTT GTC GAA TAA TTA CTC TTC GAT TTG TGA TTT CTA TCT AGA TCT GGT GTT AGT
3'Oleo_3-3

UBQ10

A G
TTC TAG TTT GTG CGA TCG AAT TTG TCG ATT AAT CTG AGT TTT TCT GAT TAA CAG GCT GGA
3' Oleo_3-3

G S L L V L S G L T L A G T V I A L T I
GGA TCT CTT CTT GTT CTC TCT GGA CTT ACT CTC GCT GGA ACT GTT ATC GCT CTC ACT ATC
3' Oleo_3-3

~~~~~  
A T P L L V I F S P V L V P A V I T I F  
GCT ACA CCT CTT CTC GTT ATC TTC TCT CCT GTT CTC GTT CCT GCT GTG ATC ACT ATC TTC  
3' Oleo\_3-3  
~~~~~

~~~~~  
L L G A G F L A S G G F G V A A L S V L  
CTT CTC GGA GCT GGA TTT CTT GCT TCT GGT GGA TTT GGA GTT GCT GCT CTC TCT GTT CTC  
3' Oleo\_3-3  
~~~~~

~~~~~  
S W I Y R Y L T G K H P P G A D C L E S  
TCT TGG ATC TAC AGA TAC CTC ACT GGA AAA CAT CCT CCA GGT GCT GAT TGT CTT GAG TCT  
3' Oleo\_3-3  
~~~~~

~~~~~  
A K T K L A S C A R E M K D R A E Q F S  
GCT AAG ACT AAG CTC GCT TCT TGT GCT AGA GAG ATG AAG GAT AGA GCA GAG CAA TTC TCT  
3' Oleo\_3-3 OCS terminator  
~~~~~

~~~~~  
C Q P V A G S Q T S \* \*  
TGT CAG CCT GTT GCT GGA TCT CAG ACT TCT TAA TGA ACA TAT GGT CCT GCT TTA ATG AGA  
OCS terminator  
~~~~~

~~~~~  
TAT GCG AGA CGC CTA TGA TCG CAT GAT ATT TGC TTT CAA TTC TGT TGT GCA CGT TGT AAA  
OCS terminator  
~~~~~

~~~~~  
AAA CCT GAG CAT GTG TAG CTC AGA TCC TTA CCG CCG GTT TCG GTT CAT TCT AAT GAA TAT  
OCS terminator  
~~~~~

~~~~~  
ATC ACC CGT TAC TAT CGT ATT TTT ATG AAT AAT ATT CTC CGT TCA ATT TAC TGA TTG TAC  
OCS terminator  
~~~~~

~~~~~  
CCT ACT ACT TAT ATG TAC AAT ATT AAA ATG AAA ACA ATA TAT TGT GCT GAA TAG GTT TAT  
OCS terminator  
~~~~~

~~~~~  
AGC GAC ATC TAT GAT AGA GCG CCA CAA TAA CAA ACA ATT GCG TTT TAT TAT TAC AAA TCC  
OCS terminator  
~~~~~

~~~~~  
AAT TTT AAA AAA AGC GGC AGA ACC GGT CAA ACC TAA AAG ACT GAT TAC ATA AAT CTT ATT  
OCS terminator  
~~~~~

~~~~~  
CAA ATT TCA AAA GGC CCC AGG GGC TAG TAT CTA CGA CAC ACC GAG CGG CGA ACT AAT AAC  
OCS terminator  
~~~~~

~~~~~  
GTT CAC TGA AGG GAA CTC CGG TTC CCC GCC GGC GCG CAT GGG TGA GAT TCC TTG AAG TTG  
OCS terminator  
~~~~~

~~~~~  
AGT ATT GGC CGT CCG CTC TAC CGA AAG TTA CGG GCA CCA TTC AAC CCG GTC CAG CAC GGC  
OCS terminator  
~~~~~

~~~~~  
GGC CGG GTA ACC GAC TTG CTG CCC CGA GAA TTA TGC AGC ATT TTT TTG GTG TAT GTG GGC  
OCS terminator  
~~~~~

~~~~~  
CCC AAA TGA AGT GCA GGT CAA ACC TTG ACA GTG ACG ACA AAT CGT TGG GCG GGT CCA GGG  
OCS terminator  
~~~~~

~~~~~  
CGA ATT TTG CGA CAA CAT GTC GAG GCT CAG CAG GAC CTG CAG GCA TGC AAG CTA GCT TAC  
~~~~~

OCS terminator

CaMV35S

TAG TGA TGC ATA TTC TAT AGT GTC ACC TAA ATC TTC GAC GAA TTA ATT CCA ATC CCA CAA

CaMV35S

AAA TCT GAG CTT AAC AGC ACA GTT GCT CCT CTC AGA GCA GAA TCG GGT ATT CAA CAC CCT

CaMV35S

CAT ATC AAC TAC TAC GTT GTG TAT AAC GGT CCA CAT GCC GGT ATA TAC GAT GAC TGG GGT

CaMV35S

TGT ACA AAG GCG GCA ACA AAC GGC GTT CCC GGA GTT GCA CAC AAG AAA TTT GCC ACT ATT

CaMV35S

ACA GAG GCA AGA GCA GCA GCT GAC GCG TAC ACA ACA AGT CAG CAA ACA GAC AGG TTG AAC

CaMV35S

TTC ATC CCC AAA GGA GAA GCT CAA CTC AAG CCC AAG AGC TTT GCT AAG GCC CTA ACA AGC

CaMV35S

CCA CCA AAG CAA AAA GCC CAC TGG CTC ACG CTA GGA ACC AAA AGG CCC AGC AGT GAT CCA

CaMV35S

GCC CCA AAA GAG ATC TCC TTT GCC CCG GAG ATT ACA ATG GAC GAT TTC CTC TAT CTT TAC

CaMV35S

GAT CTA GGA AGG AAG TTC CAA GGT CAA GGT GAC GAC ACT ATG TTC ACC ACT GAT AAT GAG

CaMV35S

AAG GTT AGC CTC TTC AAT TTC AGA AAG AAT GCT GAC CCA CAG ATG GTT AGA GAG GCC TAC

CaMV35S

GCA GCA GGT CTC ATC AAG ACG ATC TAC CCG AGT AAC AAT CTC CAG GAG ATC AAA TAC CTT

CaMV35S

CCC AAG AAG GTT AAA GAT GCA GTC AAA AGA TTC AGG ACT AAT TGC ATC AAG AAC ACA GAG

CaMV35S

AAA GAC ATA TTT CTC AAG ATC AGA AGT ACT ATT CCA GTA TGG ACG ATT CAA GGC TTG CTT

CaMV35S

CAT AAA CCA AGG CAA GTA ATA GAG ATT GGA GTC TCT AAA AAG GTA GTT CCT ACT GAA TCT

CaMV35S

AAG GCC ATG CAT GGA GTC TAA GAT TCA AAT CGA GGA TCT AAC AGA ACT CGC CGT GAA GAC

CaMV35S

TGG CGA ACA GTT CAT ACA GAG TCT TTT ACG ACT CAA TGA CAA GAA GAA AAT CTT CGT CAA

CaMV35S

CAT GGT GGA GCA CGA CAC TCT GGT CTA CTC CAA AAA TGT CAA AGA TAC AGT CTC AGA AGA

CaMV35S

CCA AAG GGC TAT TGA GAC TTT TCA ACA AAG GAT AAT TTC GGG AAA CCT CCT CGG ATT CCA
CaMV35S

TTG CCC AGC TAT CTG TCA CTT CAT CGA AAG GAC AGT AGA AAA GGA AGG TGG CTC CTA CAA
CaMV35S

ATG CCA TCA TTG CGA TAA AGG AAA GGC TAT CAT TCA AGA TCT CTC TGC CGA CAG TGG TCC
CaMV35S

CAA AGA TGG ACC CCC ACC CAC GAG GAG CAT CGT GGA AAA AGA AGA CGT TCC AAC CAC GTC
CaMV35S

TTC AAA GCA AGT GGA TTG ATG TGA CAT CTC CAC TGA CGT AAG GGA TGA CGC ACA ATC CCA
CaMV35S

CTA TCC TTC GCA AGA CCC TTC CTC TAT ATA AGG AAG TTC ATT TCA TTT GGA GAG GAC ACG
5' DGAT1 (S205A)

NcoI

M A I L D S A G V T T V
GGA TCC TTG CTC CGT TAA AAA AAA CCA TGG CTA TCC TCG ATT CTG CTG GTG TTA CTA CTG
5' DGAT1 (S205A)

T E N G G G E F V D L D R L R R R K S R
TGA CTG AGA ATG GTG GTG GAG AGT TCG TTG ATC TCG ATA GAC TCA GAA GAA GAA AGT CTA
5' DGAT1 (S205A)

UBQ10

S
GAT CTG TAA ATT TCT GTG TTC CTT ATT CTC TCA AAA TCT TCG ATT TTG TTT TCG TTC GAT
UBQ10

CCC AAT TTC GTA TAT GTT CTT TGG TTT AGA TTC TGT TAA TCT TAG ATC GAA GAC GAT TTT
UBQ10

CTG GGT TTG ATC GTT AGA TAT CAT CTT AAT TCT CGA TTA GGG TTT CAT AGA TAT CAT CCG
UBQ10

ATT TGT TCA AAT AAT TTG AGT TTT GTC GAA TAA TTA CTC TTC GAT TTG TGA TTT CTA TCT
UBQ10

AGA TCT GGT GTT AGT TTC TAG TTT GTG CGA TCG AAT TTG TCG ATT AAT CTG AGT TTT TCT
UBQ10

3' DGAT1 (S205A)

D S S N G L L L S G S D N N S P S
GAT TAA CAG GAT TCT TCT AAC GGA CTT CTC CTC TCT GGA TCT GAT AAC AAC TCT CCT TCT
3' DGAT1 (S205A)

D D V G A P A D V R D R I D S V V N D D
GAT GAT GTT GGT GCT CCT GCT GAT GTG AGA GAT AGA ATC GAT TCT GTG GTG AAC GAT GAT
3' DGAT1 (S205A)

~~~~~  
A Q G T A N L A G D N N G G G D N N G G  
GCT CAA GGT ACT GCT AAC CTC GCT GGT GAT AAT AAC GGT GGA GCT GAT AAC AAT GGT GGA  
3' DGAT1 (S205A)  
~~~~~

G R G G G E G R G N A D A T F T Y R P S
GGA AGA GGT GGA GGT GAA GGT AGA GGA AAC GCT GAT GCT ACT TTC ACT TAC AGA CCA TCT
3' DGAT1 (S205A)
~~~~~

V P A H R R A R E S P L S S D A I F K Q  
GTG CCT GCT CAT AGA AGA GCT AGA GAG TCT CCT CTC TCT TCT GAT GCT ATC TTC AAG CAG  
3' DGAT1 (S205A)  
~~~~~

S H A G L F N L C V V V L I A V N S R L
TCT CAC GCT GGA CTT TTC AAC CTC TGT GTG GTG GTT CTT ATC GCT GTG AAC TCT AGA CTC
3' DGAT1 (S205A)
~~~~~

I I E N L M K Y G W L I R T D F W F S S  
ATC ATC GAG AAC CTC ATG AAG TAC GGA TGG CTC ATC AGA ACT GAT TTC TGG TTC TCT TCT  
3' DGAT1 (S205A)  
~~~~~

R S L R D W P L F M C C I S L S I F P L
AGA TCT CTC AGA GAT TGG CCT CTT TTC ATG TGC TGC ATC TCA CTC TCA ATC TTC CCT CTC
3' DGAT1 (S205A)
~~~~~

A A F T V E K L V L Q K Y I A E P V V I  
GCT GCT TTT ACT GTT GAG AAG CTC GTG CTC CAG AAG TAT ATC GCT GAA CCT GTG GTG ATC  
3' DGAT1 (S205A)  
~~~~~

F L H I I I T M T E V L Y P V Y V T L R
TTC CTC CAC ATC ATC ATC ACT ATG ACT GAG GTT CTC TAC CCT GTT TAC GTG ACT CTC AGA
3' DGAT1 (S205A)
~~~~~

C D S A F L S G V T L M L L T C I V W L  
TGC GAT TCT GCT TTC CTC TCT GGT GTT ACT CTT ATG CTC CTC ACT TGC ATT GTG TGG CTT  
3' DGAT1 (S205A)  
~~~~~

K L V S Y A H T S Y D I R S L A N A A D
AAG CTC GTG TCT TAC GCT CAC ACT TCT TAC GAT ATC AGA TCT CTC GCT AAC GCT GCT GAT
3' DGAT1 (S205A)
~~~~~

K A N P E V S Y Y V S L K S L A Y F M V  
AAG GCT AAC CCT GAA GTG TCT TAC TAC GTG TCT CTC AAG TCT CTC GCT TAC TTC ATG GTT  
3' DGAT1 (S205A)  
~~~~~

A P T L C Y Q P S Y P R S A C I R K G W
GCT CCT ACA CTT TGT TAC CAG CCA TCT TAC CCT AGA TCT GCT TGC ATT AGA AAG GGA TGG
3' DGAT1 (S205A)
~~~~~

V A R Q F A K L V I F T G F M G F I I E  
GTG GCA AGA CAA TTC GCT AAG TTG GTG ATC TTC ACT GGA TTC ATG GGA TTC ATC ATC GAG  
3' DGAT1 (S205A)  
~~~~~

Q Y I N P I V R N S K H P L K G D L L Y
CAG TAC ATC AAC CCT ATT GTG AGA AAC TCT AAG CAC CCT CTC AAG GGT GAT CTT CTC TAC
~~~~~

3' DGAT1 (S205A)

~~~~~  
A I E R V L K L S V P N L Y V W L C M F
GCT ATC GAG AGA GTT CTT AAG CTC TCT GTG CCT AAC CTT TAT GTG TGG CTC TGC ATG TTC

3' DGAT1 (S205A)

~~~~~  
Y C F F H L W L N I L A E L L C F G D R  
TAC TGT TTC TTC CAC CTC TGG CTT AAC ATC CTT GCT GAG TTG CTT TGC TTC GGA GAT AGA

3' DGAT1 (S205A)

~~~~~  
E F Y K D W W N A K S V G D Y W R M W N
GAG TTC TAC AAG GAT TGG TGG AAC GCT AAG TCT GTT GGA GAT TAT TGG AGA ATG TGG AAC

3' DGAT1 (S205A)

~~~~~  
M P V H K W M V R H I Y F P C L R S K I  
ATG CCT GTG CAT AAG TGG ATG GTG CGT CAC ATC TAC TTC CCT TGC CTC AGA TCT AAG ATC

3' DGAT1 (S205A)

~~~~~  
P K T L A I I I A F L V S A V F H E L C
CCT AAG ACT CTC GCT ATC ATT ATC GCT TTC CTC GTG TCT GCT GTT TTC CAT GAG TTG TGT

3' DGAT1 (S205A)

~~~~~  
I A V P C R L F K L W A F L G I M F Q V  
ATC GCT GTT CCT TGC AGA CTT TTC AAG CTT TGG GCT TTC CTC GGA ATC ATG TTC CAG GTT

3' DGAT1 (S205A)

~~~~~  
P L V F I T N Y L Q E R F G S T V G N M
CCA CTC GTG TTC ATC ACT AAC TAC CTC CAA GAG AGA TTC GGA TCT ACT GTT GGA AAC ATG

3' DGAT1 (S205A)

~~~~~  
I F W F I F C I F G Q P M C V L L Y Y H  
ATT TTC TGG TTC ATT TTC TGC ATC TTC GGA CAG CCT ATG TGC GTT CTC CTC TAC TAC CAC

3' DGAT1 (S205A)

~~~~~  
D L M N R K G S M S * *
GAT CTC ATG AAC AGA AAG GGA TCT ATG TCT TAA TGA AGG ATC CAC CCA GCT TTC TTG TAC

attB2

ocs

~~~~~  
AAA GTG GTG ATG GGT TCG AAA TCG ATA AGC TTG GAT CCT CTA GAG TCC TGC TTT AAT GAG

attB2

ocs

~~~~~  
ATA TGC GAG ACG CCT ATG ATC GCA TGA TAT TTG CTT TCA ATT CTG TTG TGC ACG TTG TAA

ocs

~~~~~  
AAA ACC TGA GCA TGT GTA GCT CAG ATC CTT ACC GCC GGT TTC GGT TCA TTC TAA TGA ATA

ocs

~~~~~  
TAT CAC CCG TTA CTA TCG TAT TTT TAT GAA TAA TAT TCT CCG TTC AAT TTA CTG ATT GTA

ocs

~~~~~  
CCC TAC TAC TTA TAT GTA CAA TAT TAA AAT GAA AAC AAT ATA TTG TGC TGA ATA GGT TTA

ocs

~~~~~  
TAG CGA CAT CTA TGA TAG AGC GCC ACA ATA ACA AAC AAT TGC GTT TTA TTA TTA CAA ATC
OCS

~~~~~  
CAA TTT TAA AAA AAG CGG CAG AAC CGG TCA AAC CTA AAA GAC TGA TTA CAT AAA TCT TAT  
OCS

~~~~~  
TCA AAT TTC AAA AGG CCC CAG GGG CTA GTA TCT ACG ACA CAC CGA GCG GCG AAC TAA TAA
OCS

~~~~~  
CGT TCA CTG AAG GGA ACT CCG GTT CCC CGC CGG CGC GCA TGG GTG AGA TTC CTT GAA GTT  
OCS

~~~~~  
GAG TAT TGG CCG TCC GCT CTA CCG AAA GTT ACG GGC ACC ATT CAA CCC GGT CCA GCA CGG
OCS

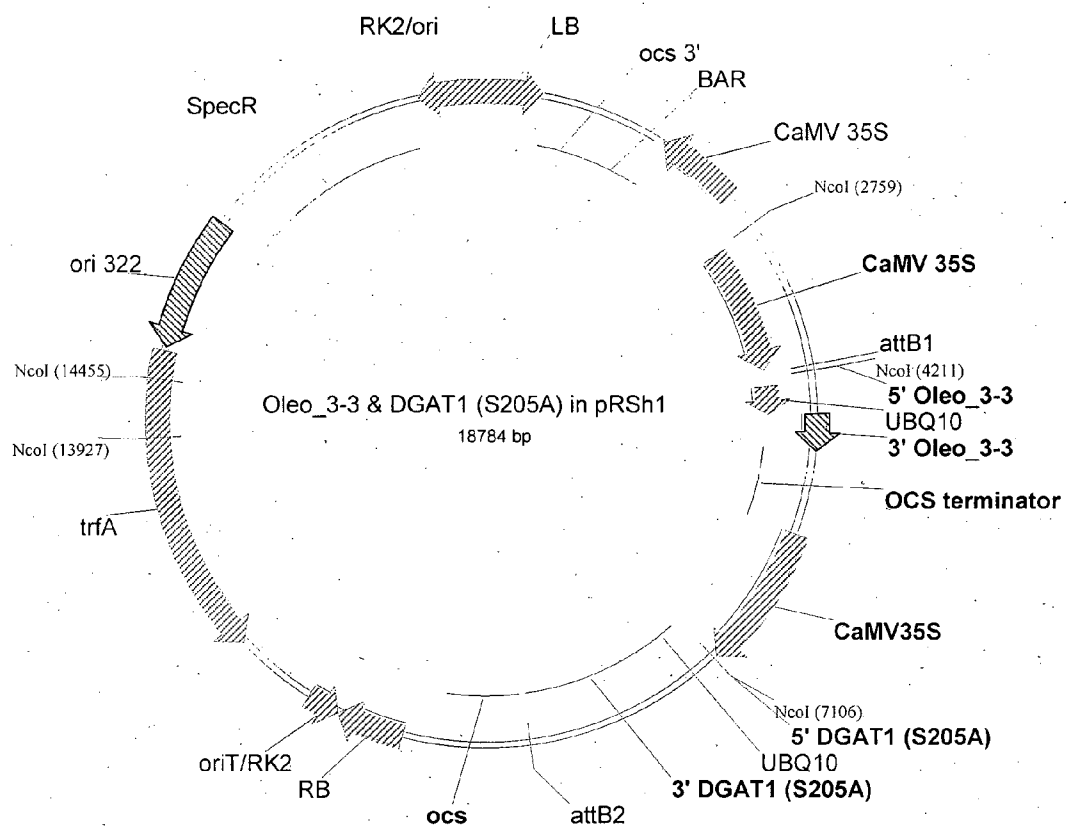
~~~~~  
CGG CCG GGT AAC CGA CTT GCT GCC CCG AGA ATT ATG CAG CAT TTT TTT GGT GTA TGT GGG  
OCS

~~~~~  
CCC CAA ATG AAG TGC AGG TCA AAC CTT GAC AGT GAC GAC AAA TCG TTG GGC GGG TCC AGG
OCS

~~~~~  
GCG AAT TTT GCG ACA ACA TGT CGA GGC TCA GCA GGA CCT GCA GGC ATG CAA GCT AGC TTA  
OCS

~~~~~  
CTA GTG ATG CAT ATT CTA TAG TGT CAC CTA AAT CT

Figure 6



42/49

Figure 7

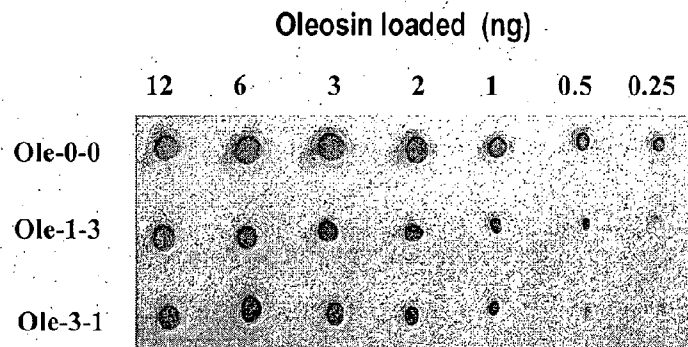


Figure 8

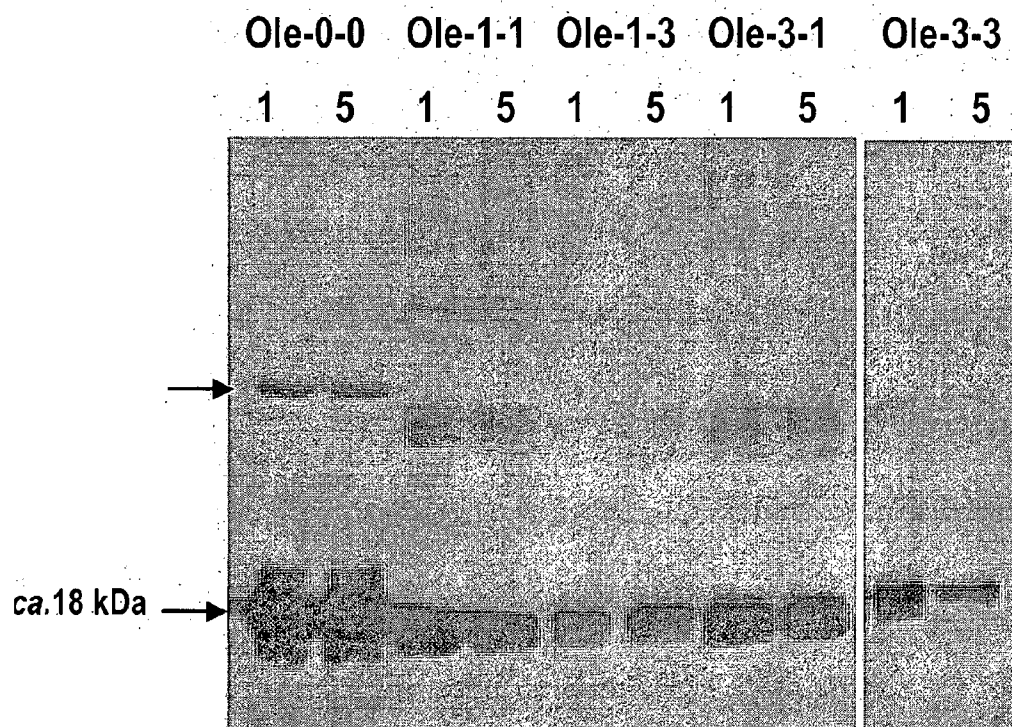


Figure 9

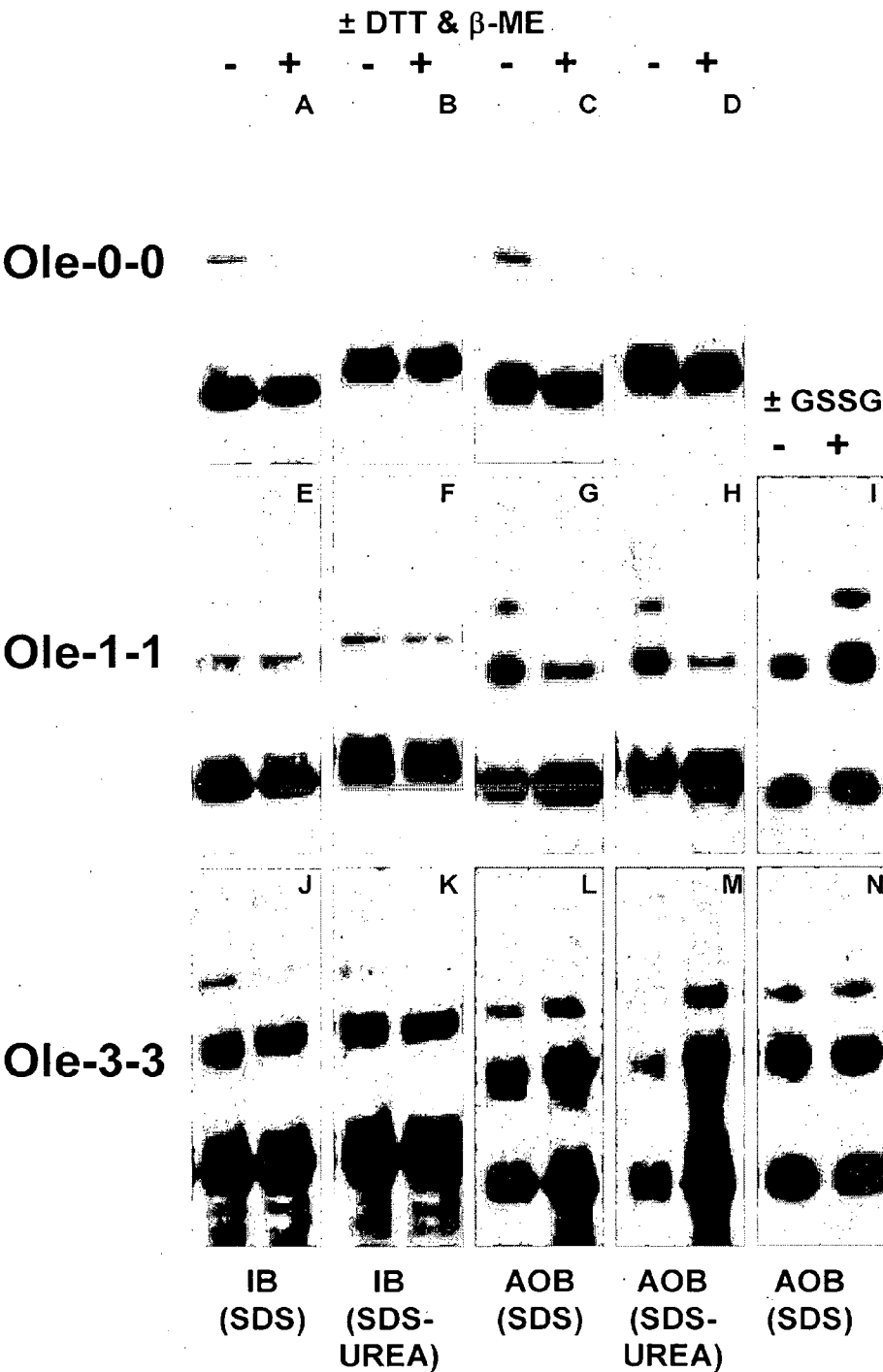


Figure 10

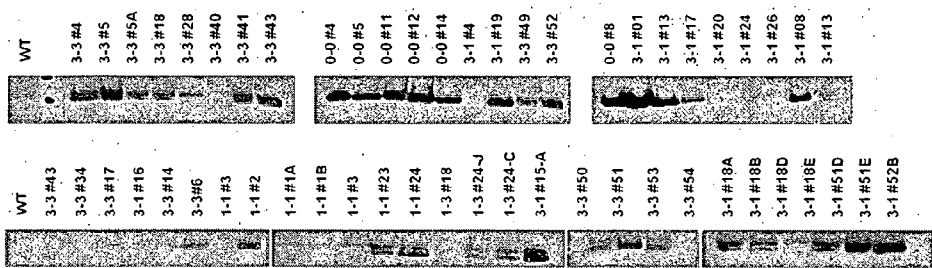


Figure 11

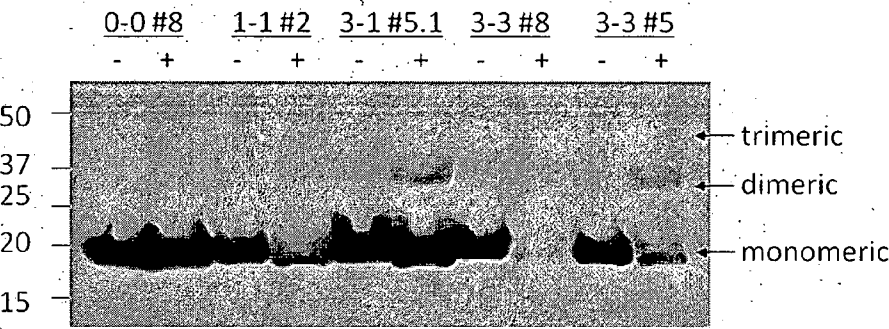


Figure 12

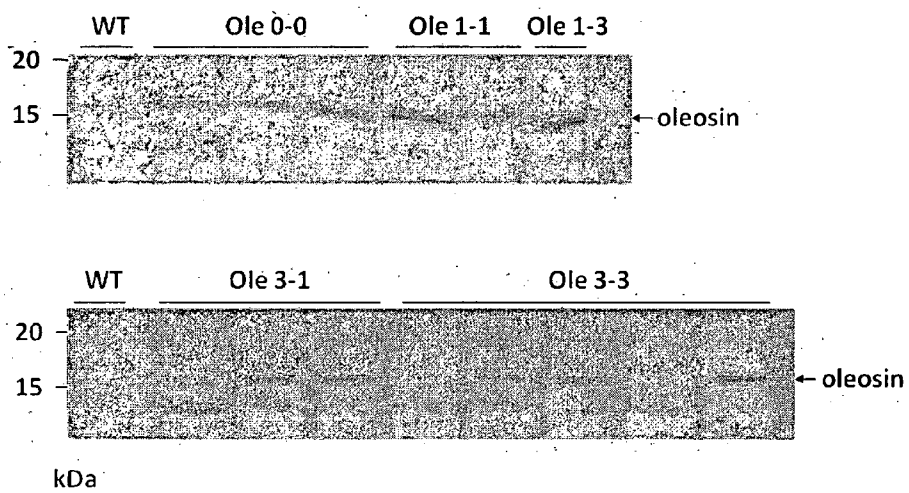


Figure 13

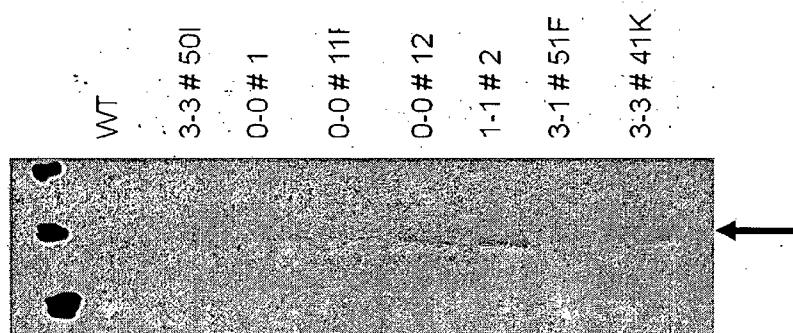


Figure 14

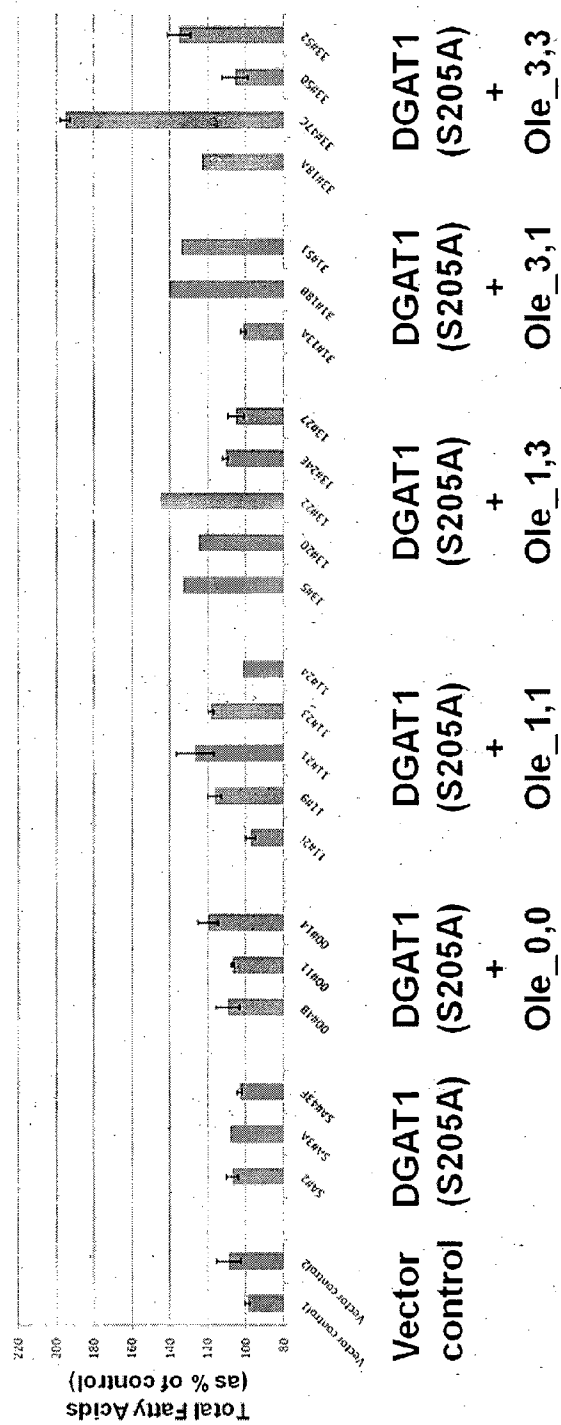


Figure 15

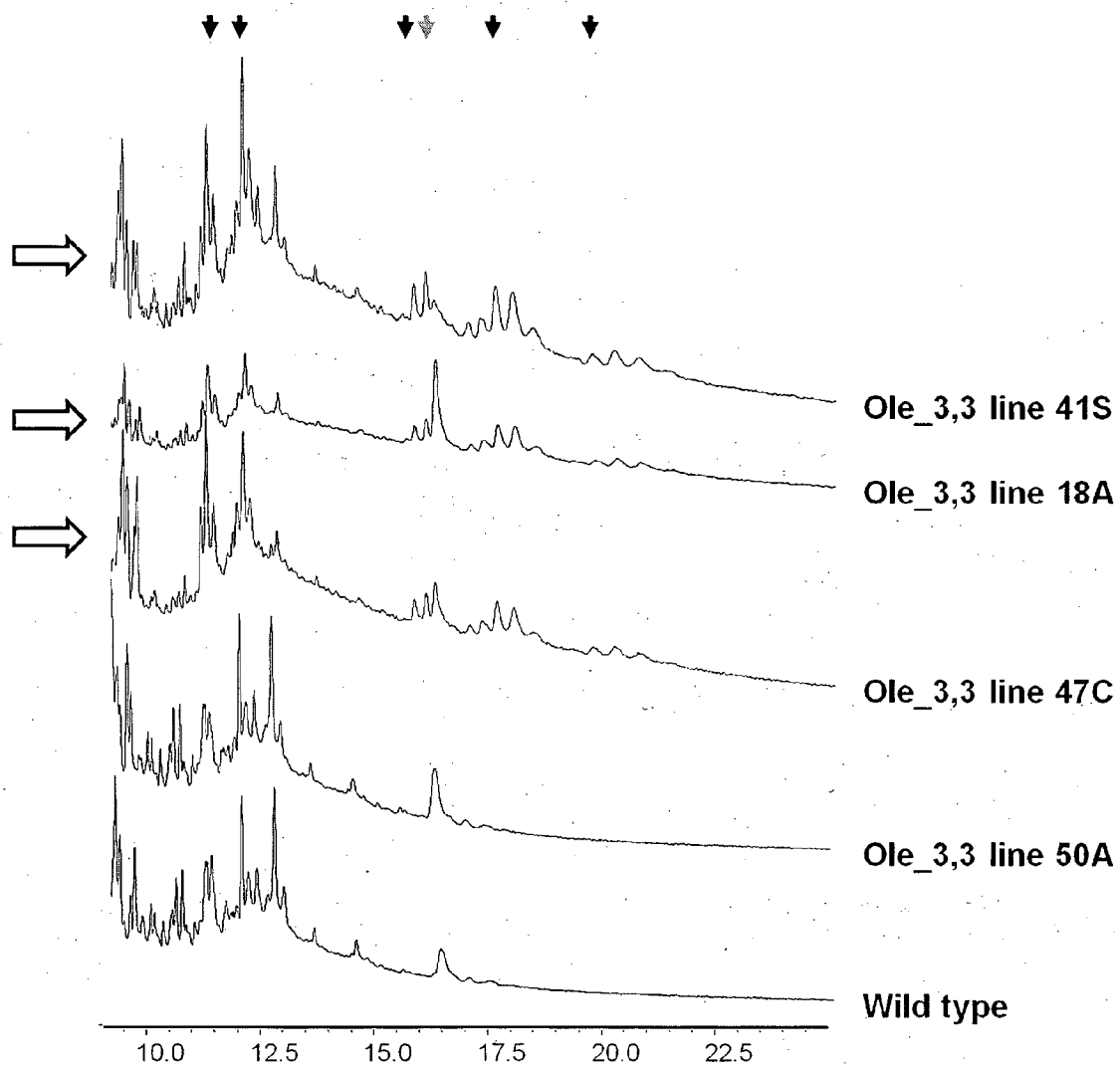


Figure 16

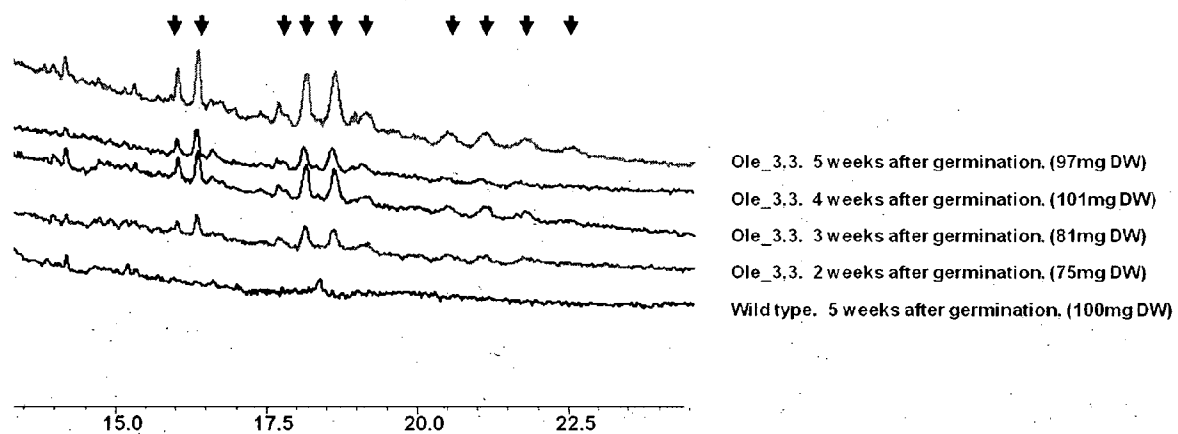


Figure 17

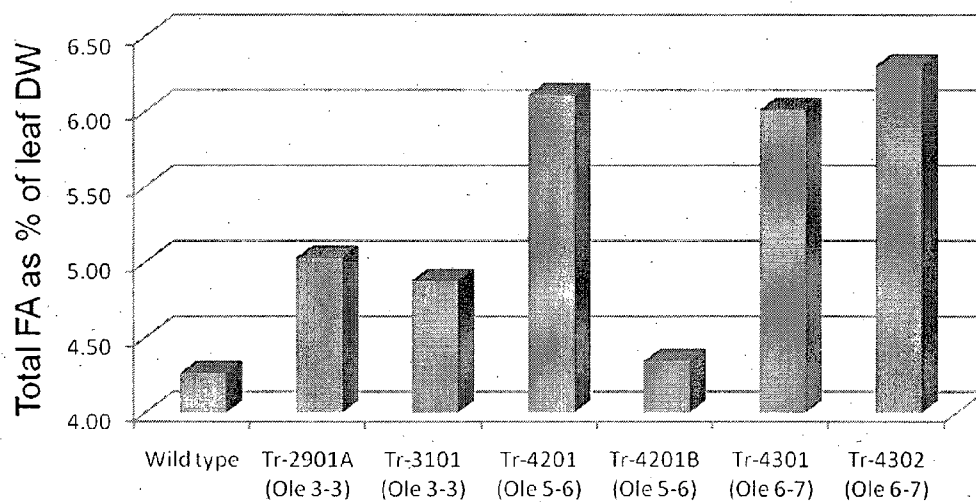


Figure 18

