



(86) Date de dépôt PCT/PCT Filing Date: 2002/09/30
(87) Date publication PCT/PCT Publication Date: 2003/04/03
(85) Entrée phase nationale/National Entry: 2004/03/29
(86) N° demande PCT/PCT Application No.: NZ 2002/000200
(87) N° publication PCT/PCT Publication No.: 2003/027141
(30) Priorité/Priority: 2001/09/28 (514547) NZ

(51) Cl.Int.⁷/Int.Cl.⁷ C07K 14/415, C12N 15/29, C12N 5/10

(71) Demandeur/Applicant:
THE HORTICULTURE AND FOOD RESEARCH
INSTITUTE OF NEW ZEALAND LIMITED, NZ

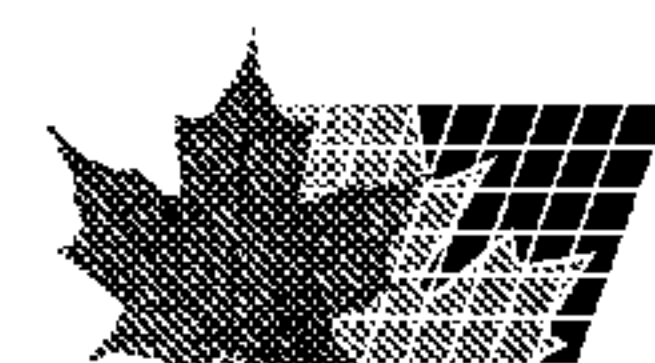
(72) Inventeurs/Inventors:
STANLEY, DUNCAN, NZ;
MACRAE, ELSPETH ANN, NZ

(74) Agent: SIM & MCBURNEY

(54) Titre : POLYPEPTIDES DE VEGETAUX ET POLYNUCLEOTIDES CODANT POUR CEUX-CI
(54) Title: PLANT POLYPEPTIDES AND POLYNUCLEOTIDES ENCODING SAME

(57) **Abrégé/Abstract:**

The present invention relates to isolated polypeptides having alpha-amylase activity and/or starch binding activity and/or plastid targeting signals and to isolated polynucleotides encoding the polypeptides. The invention also relates to DNA constructs, vectors and host cells incorporating the polynucleotide sequences, methods for modulating starch content in plants, particularly plastids, as well as modifying plastid specific starch. The applicants have identified a class of alpha-amylases with plastid targeting signals and starch binding activity.



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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
3 April 2003 (03.04.2003)

PCT

(10) International Publication Number
WO 03/027141 A1

(51) International Patent Classification⁷: **C07K 14/415**,
C12N 15/29, 5/10

(74) Agents: **CALHOUN, Douglas C A J Park** et al.; 6th Floor
Huddart Parker Building, 1 Post Office Square, Post Office
Square, PO Box 949, 6015 Wellington (NZ).

(21) International Application Number: PCT/NZ02/00200

(22) International Filing Date:
30 September 2002 (30.09.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
514547 28 September 2001 (28.09.2001) NZ

(71) Applicant (*for all designated States except US*): **THE
HORTICULTURE AND FOOD RESEARCH IN-
STITUTE OF NEW ZEALAND LIMITED** [NZ/NZ];
Corporate Office, Tennent Drive, Private Bag 11030,
Palmerston North (NZ).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, EC, EE, ES, FI, GB, GD, GE, GH, GM,
HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK,
LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX,
MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI,
SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC,
VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,
ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK,
TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, ML, MR, NE, SN, TD, TG).

(71) Applicant and

(72) Inventor: **STANLEY, Duncan**, [NZ/NZ]; 12 Phoebe
Meikle Place, Torbay, Auckland (NZ).

(72) Inventor; and

(75) Inventor/Applicant (*for US only*): **MACRAE, Elspeth,
Ann**, [NZ/NZ]; The Horticulture and Food Research Insti-
tute of New Zealand Limited, 120 Mt Albert Road, Mt Al-
bert, Auckland (NZ).

Published:

- *with international search report*
- *before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments*

*For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.*

(54) Title: PLANT POLYPEPTIDES AND POLYNUCLEOTIDES ENCODING SAME

(57) Abstract: The present invention relates to isolated polypeptides having alpha-amylase activity and/or starch binding activity and/or plastid targeting signals and to isolated polynucleotides encoding the polypeptides. The invention also relates to DNA constructs, vectors and host cells incorporating the polynucleotide sequences, methods for modulating starch content in plants, particularly plastids, as well as modifying plastid specific starch. The applicants have identified a class of alpha-amylases with plastid targeting signals and starch binding activity.



WO 03/027141 A1

PLANT POLYPEPTIDES AND POLYNUCLEOTIDES ENCODING SAME

FIELD OF THE INVENTION

5

The present invention relates to isolated polypeptides having alpha-amylase (α -amylase) activity and/or starch binding activity and/or plastid targeting signals and to isolated polynucleotides encoding the polypeptides. The invention also relates to genetic constructs, vectors and host cells incorporating the polynucleotide sequences, methods for modulating starch content in plants, particularly plastids, as well as modifying plastid specific starch.

10

BACKGROUND

15

Starch is a vital carbon storage molecule for most plants. It is insoluble and stored within the plastids of plant cells. Starch can be separated into one of two basic groups by its function. The first, transitory (or diurnal) starch is produced in photosynthetic tissues during the day, and at night it is broken down into sugars and exported to other parts of the plant. The recipient tissues can feed the sugars directly into catabolic pathways, or they can re-synthesise starch for longer-term storage. This storage starch forms the second group and is produced in a variety of organs, including roots, tubers, bark/wood and developing fruit.

20

α -Amylase (E.C. 3.2.1.1) is a starch endo-hydrolase, which cleaves α -1,4-glucan bonds within starch molecules. It is made up of three domains: domain A folds into a $(\beta/\alpha)_8$ barrel, and contains the catalytic residues of the enzyme; domain B is a large loop that protrudes from between the third β -strand and the third α -helix of domain A; domain C is located at the C-terminal end of the $(\beta/\alpha)_8$ barrel, and is made up of β -strands. The functions of domains B and C are largely unknown, although domain B has been shown to influence several isozyme-specific properties of barley α -amylases, including substrate binding, catalysis, and stability at low pH.

25

30

Much of the work on plant α -amylases has focussed on enzymes from monocotyledonous plants, particularly the cereals barley and rice. α -Amylase plays a vital role in the germination of cereal grains; it is secreted from the aleurone layer into the endosperm where it initiates starch hydrolysis. However α -amylases are not limited to tissues of germinating

seeds. α -Amylase activity has also been detected in water stressed leaves of barley, and in cultured rice cells, localised to the cell walls and amyloplasts. Molecular studies of α -amylase genes in monocots have revealed a large but well conserved family, often represented by multiple genes in each plant, e.g. at least 10 genes in rice. The proteins encoded by these genes all have N-terminal signal peptides, which would channel them into the cellular secretory pathway. No monocot α -amylases targeted to intracellular locations have been identified to date.

Dicot α -amylases have received very little attention compared to the monocot enzymes. To date, only a handful of α -amylase genes have been identified in dicots, predominantly in germinating cotyledons, most with high homology to previously characterised monocot genes. As for monocots, most dicot α -amylases possess putative signal peptides.

For the known dicotyledenous α -amylases, specific function and sub-cellular localisation has yet to be elucidated. Given the importance of starch storage in plants and plastids in particular, it would be desirable to identify α -amylases implicated to function in plastids.

It would also be desirable to identify polypeptide signal sequences directing such plastid-localised α -amylases to plastids. The nucleotide sequences encoding such plastid targeting signals could be utilised to direct chimeric proteins, containing the plastid targeting signals, to the plastids of transgenic plants. This compartmentalisation could avoid any toxic effects of expressed protein, on the contents of the cell outside the plastids. Additionally plastid localisation could avoid the deleterious effects of cytoplasmic factors on the expressed protein. It would also be advantageous to target polypeptides intended to interact with plastid factors, to plastids.

Further, it would be desirable to identify polypeptide sequences implicated in starch binding in plants. Chimeric recombinant proteins including such starch binding sequences could be produced. Such recombinant proteins could have high value in industrial processes. For example they could be used to modify industrial materials that use starch or starch derivatives as their polymeric base, such as biodegradable films.

The applicants have now identified and isolated from apple a polynucleotide encoding a novel α -amylase polypeptide which contains a plastid targeting signal. It is broadly towards this

polynucleotide, to its homologs and to the modulation of its expression/function within plants that the present invention is directed.

5 The applicants have also identified, for the first time, in plant α -amylases, peptide motifs implicated in plastid targeting of α -amylases. The invention also relates utilisation of this sequence and its homologues, to facilitate plastid targeting of chimeric protein in transgenic plants.

10 The applicants have further identified for the first time in plant α -amylases, polypeptide motifs implicated in starch binding of α -amylases. The invention further relates to utilization of these peptide sequences, to produce recombinant chimeric proteins with starch binding properties.

15 SUMMARY OF THE INVENTION

In a first aspect, the present invention provides an isolated polypeptide, which encodes a plastid α -amylase.

In a further aspect the present invention provides an isolated polypeptide having a mature sequence derived from a plastid alpha-amylase with an N-terminal extension comprising both a plastid targeting polypeptide and a starch-binding domain.

In a further aspect the present invention provides an isolated polypeptide with an N-terminal extension which comprises:

- a) a plastid targeted polypeptide of an amino acid sequence selected from SEQ ID NO: 7, SEQ ID NO: 8, residues 1-70 of SEQ ID NO: 10 and residues 1-53 of SEQ ID NO: 53; or a functionally equivalent variant thereof, and
- b) at least one repeat of the defined polypeptide motif pair:

Motif 1 y HW (G/A)y[X]₆₋₉ WXXP[X]₃₋₅PXX(T/S)

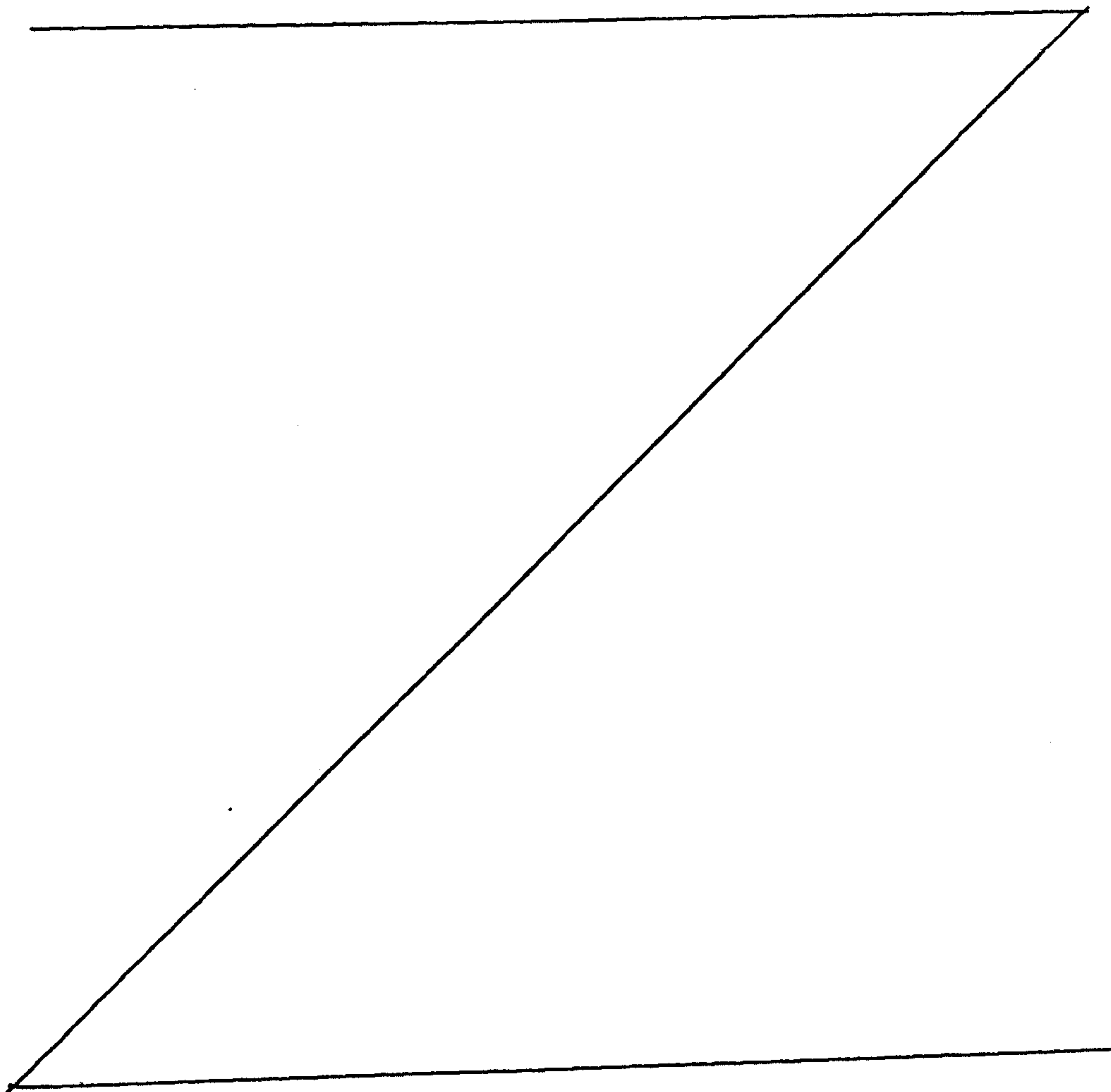
Motif 2 F(V/L)y[X]₅₋₈ W[X]₆₋₈(D/N)F

In a further aspect, the present invention provides an isolated polypeptide having the *Mdamy10* amino acid sequence of SEQ ID NO:2 or a functionally equivalent variant thereof. The invention also provides the isolated mature polypeptide.

In a further aspect, the present invention provides an isolated polynucleotide which encodes a plastid targeted α -amylase polypeptide having the *Mdamy10* amino acid sequence of SEQ ID NO:2 or a functionally equivalent variant thereof.

Preferably, said polynucleotides comprise part or all of the nucleotide sequence of SEQ ID NO:1.

In one embodiment, there is provided a polynucleotide sequence of the invention, which encodes a novel N-terminal domain of an α -amylase, and comprises the sequence of SEQ ID NO:3.



The invention also provides a polynucleotide sequence of the invention which encodes the N-terminal domain comprising of SEQ ID NO:4 or a functionally equivalent variant thereof.

The invention also provides a polynucleotide sequence which encodes a novel starch binding domain polypeptide of SEQ ID NO:5 or a functionally equivalent variant thereof.

The invention also provides a polynucleotide sequence of SEQ ID NO:6.

Also provided is a polynucleotide sequence which encodes a plastid targeting polypeptide selected from SEQ ID NO:7, SEQ ID NO:8, residues 1-70 of SEQ ID NO:10 and residues 1-53 of SEQ ID NO:53 or a functionally equivalent variant thereof.

The invention also provides a polynucleotide sequence coding for a polypeptide which comprises at least one repeat, preferably two, of the defined polypeptide motif pair. The motif pair can be defined by aligning of the N-terminal half of the family three α -amylases, namely *Mdamy10*, *Mdamy11*, *Atamy3*, *Osamy10* (from rice, *Oryza sativa*), and Fragment 1 from kiwifruit (*Actinidia chinensis*)(Figure 8):

Motif 1: yHWGV[X]₇₋₁₀W(D/E)(Q/I)P(P)[X]₃₋₄P(P)[X]₈A(I/L)XTXL

Motif 2: FV(F/L/V)K[X]₂E[X]₂₋₃W[X]₄₋₆GXDF

or functionally equivalent variants thereof.

(In this notation, capital letters represent conserved amino acids, whilst letters in parentheses represent partly conserved amino acids; y represents a hydrophobic residue; X represents any amino acid; and [X]₄₋₆ represents a run of 4 to 6 unspecified amino acids).

The polynucleotide is preferably DNA.

The motifs can be more loosely defined if sequences from other putative starch binding proteins are included. The sequences added were R1 protein from potato, its homolog from *Arabidopsis*, Sex1, and a putative starch branching enzyme (SBE-like) from *Arabidopsis* (Figure 9). The resulting motifs are:

Motif 1: yHW(G/A)y[X]₆₋₉WXXP[X]₃₋₅PXX(T/S)

Motif 2: F(V/L)y[X]₅₋₈W[X]₆₋₈(D/N)F

or functionally equivalent variants thereof. The same notation is used as above.

The polynucleotide is preferably DNA.

The invention further provides a genetic construct, which includes a polynucleotide as defined above.

- 5 More particularly, the invention provides a genetic construct comprising in the 5'-3' direction:
- (a) a promoter sequence;
 - (b) an open reading frame polynucleotide coding for a polypeptide of the invention or a functionally equivalent variant thereof; and
 - 10 (c) a termination sequence.

Preferably, the polypeptide comprises an amino acid sequence selected from SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7 and SEQ ID NO:8.

- 15 In one embodiment, the open reading frame is in a sense orientation.

In an alternate embodiment, the open reading frame is in an anti-sense orientation.

- In still a further embodiment, the invention provides a genetic construct comprising, in the
- 20 5'-3' direction:
- (a) a promoter sequence;
 - (b) a non-coding region of, a polynucleotide which encodes a polypeptide of the invention or a functionally equivalent variant thereof; and
 - (c) a termination sequence.

25

Preferably, the polypeptide has an amino acid sequence selected from SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7 and SEQ ID NO:8.

Once again, the non-coding region can be in a sense or anti-sense orientation.

30

In yet a further embodiment, the invention provides a genetic construct comprising, in the 5'-3' direction:

- (a) a promoter sequence;

- (b) a polynucleotide comprising a polynucleotide sequence complementary to at least part of a sequence coding for a polypeptide of the invention, or a functionally equivalent variant of either; and
- (c) a termination sequence.

5

In one embodiment, the polynucleotide of the genetic construct in (b), includes an inverted repeat, such that a hairpin structure can be formed from the transcript.

Preferably, in each embodiment, the genetic construct further includes a marker for
10 identification of transformed cells.

In a further aspect, the invention provides a host cell, which includes a genetic construct as defined above.

15 In still a further aspect, the invention provides a transgenic plant cell which includes a genetic construct as defined above, as well as a transgenic plant comprising such cells.

In a yet further aspect, the invention provides a plant which has been genetically modified to alter the expression and/or activity of a polypeptide of the sequence in SEQ ID NO:2 or a
20 functionally equivalent variant thereof.

In a further aspect, the invention provides a plant which contains a genetic construct comprising a polynucleotide encoding a polypeptide having the *Mdamy10* SEQ ID NO: 2 or SEQ ID NO:4 or a functionally equivalent variant thereof and in which expression and/or
25 activity of said polypeptide within said plant has been disrupted.

Such a plant will accumulate starch in the plastid and in the organ where starch is stored (e.g. leaves, tubers, roots, bark/wood, fruit) moreover the starch structure and composition will be altered. In addition primary metabolism will be altered in the transformed tissue and the plant
30 will have altered carbohydrate transport between the various organs.

In a further aspect, the invention provides a plant which has been genetically modified such that it does not functionally express a polypeptide selected from SEQ ID NO:4 and SEQ ID NO:5.

35

In such a plant starch binding capacity of the polypeptide is reduced which means proteins affected are unable to function in starch metabolism.

5

In one form, functional expression of said polypeptide encoded by the polynucleotide is disrupted directly.

10

In another form, functional expression of said polypeptide encoded by polynucleotide is disrupted indirectly, such as through disrupting functional expression of the polypeptide encoded by said polynucleotide.

15

As used herein "plant" means gymnosperm, angiosperm, monocotyledenous or dicotyledenous plants, plant parts such as leaves, roots, flowers, fruit, bark/wood, tubers, cuttings, seeds, tissue cultures, cell cultures and plant cells but is not limited thereto.

20

As used herein, "functional expression" of said polypeptide refers to the amount of the polypeptide which is expressed and functional within the plant. For example, a plant which does not functionally express a polypeptide can mean either that there is no expression of that polypeptide at all, or that the polypeptide is expressed but no longer performs its previous function.

25

Disruption of expression and/or activity may be by mutation (such as frameshift, deletion, insertion or knockout mutations) of the polynucleotide itself or of its regulatory elements, or by down-regulation (such as antisense or co-suppression) or any other method known to those skilled in the art by which aberrant or reduced expression of the gene may be achieved. Disruption may therefore be specifically caused by down-regulation of expression of *Mdamy10*.

30

It is appreciated that experiments designed to decrease expression levels of polypeptides of the invention, may result in a range of expression levels in different transgenic plants. The different levels of expression will result in different levels of α -amylase activity, any of which may provide useful plants.

The invention also provides a method for modulating the starch content of a plant, the method comprising increasing or decreasing expression and/or activity of the polypeptide selected from SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7 and SEQ ID NO:8 by genetic modification to alter the expression of a gene encoding a plastid targeted α -amylase.

5

Starch metabolism may be altered by introducing into a plant a genetic construct of the invention and expressing the polypeptide in the plant.

10 The invention also broadly relates to the use of the polynucleotides and polypeptides of the invention to modulate starch content of organisms including plants.

Also specifically contemplated is the use of the N-terminal and starch binding domain and/or specific starch binding motifs in the production of recombinantly expressed chimeric
15 polypeptides containing such starch binding domains and or motifs. Such recombinant chimeric polypeptides could possess the ability to bind starch.

Also specifically contemplated is the use of the polynucleotide sequences encoding the plastid targeting motifs of the invention, to prepare chimeric genetic constructs, which direct the
20 targeting of transgenically expressed polypeptides to the plastids of transgenic plants.

While the invention is broadly defined as above, those persons skilled in the art will appreciate that it is not limited thereto and that it also includes embodiments of which the following description provides examples.

25

BRIEF DESCRIPTION OF THE DRAWINGS

In addition, the present invention will be better understood from reference to the
30 accompanying drawings in which:

Figure 1 shows a comparative alignment of Family three α -amylase sequences with apple α -amylase 8 (*Mdamy8*). *Mdamy8* is a cytosol-targeted α -amylase. Each bar is shaded to distinguish the four protein domains encoded by the sequences: A, B and C represent the

structural domains A, B and C, which are found throughout all α -amylases; Novel domain refers to the N-terminal domain found only in Family three α -amylases. *Mdamy8*, *Atamy3*, *Mdamy10*, and *Osamy10* all represent full-length sequences. *Mdamy11* and sequences from Kiwifruit, Blueberry, Coffee, Cotton, and Rose, all represent partial sequences.

5

Figure 2 shows the nucleotide sequence of *Mdamy10*. The nucleotides encoding the plastid targeting peptide (nucleotides 55 to 237) are shown in bold, while the nucleotides encoding the novel starch binding domain are underlined (238 to 1503).

10 Figure 3 shows the amino acid sequence of *Mdamy10*. The plastid targeting peptide (amino acids 1 to 61) is shown in bold, the novel starch binding domain is underlined (62 to 483).

15 Figure 4 shows the nucleotide sequence of *Atamy3*. The nucleotides encoding the plastid targeting peptide (nucleotides 1 to 165) are shown in bold, while the nucleotides encoding the novel starch binding domain are underlined (166 to 1410).

Figure 5 shows the amino acid sequence of *Atamy3*. The plastid targeting peptide (amino acids 1 to 55) is shown in bold, the novel starch binding domain is underlined (56 to 470).

20 Figure 6 shows the nucleotide sequence of *Osamy10*. The nucleotides encoding the plastid targeting peptide (nucleotides 1 to 159) are shown in bold, while the nucleotides encoding the novel starch binding domain are underlined (160 to 1371).

25 Figure 7 shows the amino acid sequence of *Osamy10*. The plastid targeting peptide (amino acids 1 to 53) is shown in bold, the novel starch binding domain is underlined (54 to 457).

30 Figure 8 shows an alignment of polypeptide sequences from the starch binding domains of Family three α -amylases. Sequence titles are shown to the left of the sequences themselves. The kiwifruit sequence refers to sequence 1 (SEQ ID NO 21). The letter 'a' indicates that the sequence is the first (closest to the N-terminus) repeat of a motif pair within the sequence; 'b' indicates the second (closest to the C-terminus) repeat of a motif pair. Shading indicates sequence conservation: Black = fully conserved, white letters on gray background = highly conserved, black letters on gray background = moderately conserved.

Figure 9 shows an alignment of polypeptide sequences from the starch binding domains of Family three α -amylases, plus R1 protein from potato (Genbank - CAA70725), its homolog from *Arabidopsis*, Sex1 (Genbank - AAG47821), and a putative starch branching enzyme (SBE-like) from *Arabidopsis* (Genbank - BAB02827). Sequence titles are shown to the left of the sequences themselves. The kiwifruit sequence refers to sequence 1 (SEQ ID NO 21). The letter 'a' indicates that the sequence is the first (closest to the N-terminus) repeat of a motif pair within the sequence; 'b' indicates the second (closest to the C-terminus) repeat of a motif pair. Shading indicates sequence conservation: Black = fully conserved, white letters on gray background = highly conserved, black letters on gray background = moderately conserved.

Figure 10 shows results of PCR with primers NewUNIf2 and NewUNIr2 on genomic DNA from a number of plant species: Lane 1; pine. Lane 2; potato. Lane 3; onion. Lane 4; rose. Lane 5; olive. Lane 6; coffee. Lane 7; banana. Lane 8; mango. Lane 9; cotton. Lane 10; 1kb+ DNA ladder (Bio-Rad Laboratories). Indicative sizes are shown to the right. Samples were electrophoresed through a 0.8% agarose gel and stained with ethidium bromide.

Figure 11 shows fluorescence microscopy images of GFP fusion genetic constructs expressed in *N. benthamiana* epidermal cells. Top left, GFP alone (pDS-GFP-ART27), using GFP plant filter set; bar represents 50 μ m. Top right, cTP-GFP (pcTP-ART27), using UV filter set; bar represents 50 μ m. Bottom left, cTP-GFP, using GFP plant filter set; bar represents 50 μ m. Bottom right, cTP-GFP, using GFP plant filter set; bar represents 10 μ m.

Fig 12 Shows SDS-PAGE (A) and western blot (B) of protein from IPTG induced *E. coli* containing pET-30a-based GFP constructs. The western was probed with anti-GFP antibodies. Lanes 1-4 are insoluble protein fractions. Lanes 6-9 are soluble protein fractions. Lanes 1 and 6 – pET-30a; Lanes 2 and 7 – pGFP-ET-30b; Lanes 3 and 8 – pNterm-ET-30b; Lanes 4 and 9 – pSBD-ET-30a; Lane 5 – Prestained Broad Range SDS-PAGE Standard (Bio-Rad) – sizes shown at right of figure.

Figure 13 shows protein samples on amylopectin-containing gels that have been stained with iodine. Gel A: Lane 1 contains crude protein extract from *E. coli* expressing pET-30a. Lane 2 contains crude protein extract from *E. coli* expressing pAmy10-ET-30a. Lane 3 contains crude pET-30a extract desalted in binding buffer. Lane 4 contains crude pAmy10-ET-30a

extract desalted in binding buffer. Gel B: Lanes 5, 7 and 9 contain nickel-purified pET-30a protein, from consecutive 2.5mL fractions. Lanes 6, 8 and 10 contain nickel-purified pAmy10-ET-30a protein, from consecutive 2.5mL fractions. Lane 11 contains crude protein extract from *E. coli* expressing pAmy10-ET-30a.

5

Figure 14 shows a starch transfer gel that has been stained with iodine. Lane 1 contains crude protein extract from *E. coli* expressing pET-30a only. Lane 2; crude protein extract from *E. coli* expressing pAmy10-ET-30a. Lane 3; protein extracted from apple leaves. Lane 4; protein extracted from *Arabidopsis* leaves. Lane 5; crude protein extract from *E. coli* expressing pAmy10-ET-30a.

10

Figure 15 shows SDS-PAGE (A) and western blot (B) of protein from IPTG induced *E. coli* containing pET-30a and pAmy10-ET-30a plasmids. The western was probed with anti-His₆ antibodies. Lane 1 – crude pET-30a; Lane 2 – crude pAmy10-ET-30a; Lane 3 – desalted pET-30a; Lane 4 – desalted pAmy10-ET-30a; Lane 5 – Pre-stained Broad Range SDS-PAGE Standard (Bio-Rad) – sizes shown at right of figure; Lane 6 - pAmy10-ET-30a, purification fraction 1; Lane 7 - pET-30a, purification fraction 2; Lane 8 - pAmy10-ET-30a, purification fraction 2; Lane 9 - pAmy10-ET-30a, purification fraction 3 (part of the lane was removed prior to semi-dry blotting, and so does not appear complete on the blot (B)).

20

Figure 16 shows RT-PCR of *Arabidopsis* α -amylase transcripts from ten different tissues. PCR amplification was carried out for thirty cycles, using primers designed to amplify part of each *Arabidopsis* α -amylase (Atamy1, 2 or 3). Tissues: 1: Imbibed seeds, 2: Seeds with emerging cotyledons, 3: Cotyledons only from (2), 4: Whole seedling (2 leaves), 5: Leaves only from (4), 6: Roots, 7: Full rosette plant, 8: Leaves from growing stem, 9: Young seed pods, 10: Senescing plant (minus seed pods).

25

DESCRIPTION OF THE INVENTION

As broadly outlined above, the applicants have identified novel α -amylases targeted to the plastid of a plant. In a preferred embodiment the plant is a fruiting plant. Polynucleotides coding for the novel polypeptides are also provided. The specific polypeptides and polynucleotides are from the plants *Malus x domestica*, *Arabidopsis thaliana* and *Oryza sativa*.

The amino acid sequence of one polypeptide, *Mdamy10*, and that of the polynucleotide sequence encoding it are given in Figures 3 and 2 respectively. It will however be appreciated that the invention is not restricted only to the polypeptide/polynucleotide having the specific amino acid/nucleotide sequence given in Figures 3 and 2. Instead, the invention also extends to functionally equivalent variants of the polypeptide/polynucleotide of Figures 3 and 2.

The term "polynucleotide(s)" as used herein means a single or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases and includes DNA and corresponding RNA molecules, including hnRNA and mRNA molecules, both sense and anti-sense strands, and comprehends cDNA, genomic DNA and recombinant DNA, as well as wholly or partially synthesized polynucleotides. An hnRNA molecule contains introns and corresponds to a DNA molecule in a generally one-to-one manner. An mRNA molecule corresponds to an hnRNA and DNA molecule from which the introns have been excised. A polynucleotide may consist of an entire gene, or any portion thereof. Operable anti-sense polynucleotides may comprise a fragment of the corresponding polynucleotide, and the definition of "polynucleotide" therefore includes all such operable anti-sense fragments.

The term 'polypeptide(s)' as used herein includes peptides, polypeptides and proteins.

The phrase "functionally equivalent variants" recognises that it is possible to vary the amino acid/nucleotide sequence of a polypeptide while retaining substantially equivalent functionality. For example, a polypeptide can be considered a functional equivalent of another polypeptide for a specific function if the equivalent polypeptide is immunologically cross-reactive with and has at least substantially the same function as the original polypeptide. The equivalent can be, for example, a fragment of the polypeptide, a fusion of the polypeptide with another polypeptide or carrier, or a fusion of a fragment with additional amino acids.

Variant polynucleotide sequences also include equivalent sequences, which vary in size, composition, position and number of introns, as well as size and composition of untranslated terminal regions.

5

Functionally equivalent polynucleotides are those encoding functionally equivalent polypeptides.

The polynucleotide sequence encoding important N-terminal polypeptide region of *Mdamy10* is shown in Figure 2. The corresponding N-terminal amino acid sequence from *Mdamy10* is shown in Figure 3. The large N-terminal extension (approximately 460-480 amino acids) was originally found in *Mdamy10* and a homologue from *Arabidopsis*, *Atamy3* (GenBank Accession No. A7050398; Figures 4 and 5). A further plastid-targeted sequence was identified by the applicants from apple (*Mdamy11*), and then later from rice (*Osamy10*; Figures 6 and 7). These N-terminal extensions include plastid-targeting peptides of 61 amino acids in *Mdamy10* (SEQ ID NO:2 and SEQ ID NO:7), 70 amino acids in *Mdamy11* (SEQ ID NO:10), 55 amino acids in *Atamy3* (SEQ ID NO:17 and SEQ ID NO:8) and 53 amino acids in *Osamy10* (SEQ ID NO:53).

The N-terminal regions have also been found to include a potential starch binding region (SEQ ID NO:5), and potential specific starch binding motifs as set out above. The remaining C-terminal regions (the last 420 amino acids in *Mdamy10*, the last 416 in *Atamy3*, and *Osamy10*) include the α -amylase region. *Mdamy11* sequences contain most of a starch binding domain, in amino acids 71 to 243 of the 5' polypeptide fragment (SEQ ID NO:10), which includes a complete pair of starch binding motifs, and amino acids 1 to 107 of the central fragment (SEQ ID NO:12), which includes part of a starch binding motif 2. Amino acids 108 to 172 of the central fragment and 1 to 142 of the 3' fragment (SEQ ID NO:14) all constitute part of an α -amylase region.

The applicants have also determined that *Mdamy10*, *Mdamy11*, *Atamy3*, and *Osamy10* form a distinct family of α -amylases and have named these Family three α -amylases. This was determined not only by their plastid targeting but also based on the number and distribution of their introns. *Atamy3* and *Osamy10* have 12 introns that interrupt their coding sequences; their full genomic sequences, including introns, are SEQ ID NO:55 and SEQ ID NO:54,

respectively. Six of their introns are within the α -amylase coding region of the gene (the 3' half) as shown in Table 1 below.

Table 1:

	1 ^a	2	3	4	5	6	7	8
	96-0	138-2	157-0	195-0	231-1	342-0	501-0	544-1
<i>Hvamy2-2</i>	n/a	n/a	n/a	n/a	n/a	n/a	+	–
<i>AmyVm1</i>	n/a	n/a	n/a	n/a	n/a	n/a	+	+
<i>Mdamy10</i>	?	?	?	?	?	?	?	?
<i>Mdamy11</i>	?	?	+	?	?	?	?	?
<i>Atamy3</i>	+	+	+	+	+	+	–	–
<i>Osamy10</i>	+	+	+	+	+	+	–	–

	9	10	11	12	13	14	15
	556-2	620-0	655-2	729-0	780-0	806-0	836-0
<i>Hvamy2-2</i>	–	–	–	–	–	+	–
<i>AmyVm1</i>	–	–	–	–	–	+	–
<i>Mdamy10</i>	?	?	+	+	+	?	?
<i>Mdamy11</i>	?	?	+	+	+	?	?
<i>Atamy3</i>	+	+	+	+	+	–	+
<i>Osamy10</i>	+	+	+	+	+	–	+

5

10

15

The distribution of introns in Family one and Family three α -amylase genes. The codon numbers correspond to and are equivalent to the amino acid sequence of *Atamy3* (Fig. 5). The presence of an intron is represented by a + sign and its absence by a – sign.. The regions of *Mdamy10* and *Mdamy11* for which there is no genomic DNA sequence is marked ?. The region from amino acids 1-470 of *Atamy3* has no equivalent in *Hvamy2-2* and *amyVm1*, these regions have been marked n/a. The applicants have defined a total of 15 distinct intron positions amongst the Family one and Family three genes, they are numbered from the 5' of the *Atamy3* gene sequence and defined by codon number and phase. For introns separating triplets after the first base, the codon number is followed by –1. For introns separating triplets after the second base, the codon number is followed by –2. For introns falling between triplets the number of the codon located 3' of the intron is given, followed by –0. GenBank Accession numbers: *Hvamy2-2* (AAA98790), *AmyVm1* (CAA37217).

All three α -amylases have short 5' untranslated regions (UTRs) between 35 and 54 base pairs. However, *Mdamy10* has a very long 3' UTR, of up to 557 base pairs.

- 5 Table 2 provides information on the characterisation of the UTRs of *Mdamy10*, *Mdamy11* and *Atamy3*.

Table 2:

	5' UTR Length (Max / Min) (bp)	3' UTR Length (Max / Min) (bp)	No of polyadenylation sites detected
<i>Mdamy10</i>	54	557 / 428	4
<i>Mdamy11</i>	35	98 ^a	1
<i>Atamy3</i>	46	229	Unknown

- 10 ^a Refers to the 3' UTR of the mis-spliced, truncated *Mdamy11* transcript.

The unique intron structure which characterises this α -amylase family is also believed to affect expression of the α -amylases.

- 15 It will be understood that a variety of substitutions of amino acids is possible while preserving the structure responsible for activity of the polypeptides. Conservative substitutions are described in the patent literature, as for example, in United States Patent No 5,264,558 or 5,487,983. It is thus expected, for example, that interchange among non-polar aliphatic neutral amino acids, glycine, alanine, proline, valine and isoleucine, would be possible.
- 20 Likewise, substitutions among the polar aliphatic neutral amino acids, serine, threonine, methionine, asparagine and glutamine could be made. Substitutions among the charged acidic amino acids, aspartic acid and glutamic acid, could probably be made, as could substitutions among the charged basic amino acids, lysine and arginine. Substitutions among the aromatic amino acids, including phenylalanine, histidine, tryptophan and tyrosine are also possible.
- 25 Such substitutions and interchanges are well known to those skilled in the art.

Equally, nucleotide sequences encoding a particular product can vary significantly simply due to the degeneracy of the nucleic acid code.

A polynucleotide or polypeptide sequence may be aligned, and percentage of identical nucleotides in a specified region may be determined against another sequence, using computer algorithms that are publicly available. Two exemplary algorithms for aligning and identifying the similarity of polynucleotide sequences are the BLASTN and FASTA algorithms. The similarity of polypeptide sequences may be examined using the BLASTP algorithm. Both the BLASTN and BLASTP software are available on the NCBI anonymous FTP server (<ftp://ncbi.nlm.nih.gov>) under `/blast/executables/`. The BLASTN algorithm version 2.0.4 [Feb-24-1998], set to the default parameters described in the documentation of variants according to the present invention. The use of the BLAST family of algorithms, including BLASTN and BLASTP, is described at NCBI's website at URL <http://www.ncbi.nlm.nih.gov/BLAST/newblast.html> and in the publication of Altschul, Stephen F., *et al.* (1997). The computer algorithm FASTA is available on the Internet at the ftp site <ftp://ftp.virginia.edu/pub/fasta/>. Version 2.0u4, February 1996, set to the default parameters described in the documentation and distributed with the algorithm, is also preferred for use in the determination of variants according to the present invention. The use of the FASTA algorithm is described in (Pearson and Lipman 1988, Pearson 1990).

The following running parameters are preferred for determination of alignments and similarities using BLASTN that contribute to E values (as discussed below) and percentage identity: Unix running command: `blastall -p blastn -d embldb -e 10 -G 1 -E 1 -r 2 -v 50 -b 50 -I queryseq -o results`; and parameter default values:

`-p` Program Name [String]
`-d` Database [String]
`-e` Expectation value (E) [Real]
`-G` Cost to open a gap (zero invokes default behaviour) [Integer]
`-E` Cost to extend a gap (zero invokes default behaviour) [Integer]
`-r` Reward for a nucleotide match (blastn only) [Integer]
`-v` Number of one-line descriptions (V) [Integer]
`-b` Number of alignments to show (B) [Integer]
`-i` Query File [File In]
`-o` BLAST report Output File [File Out] Optional

For BLASTP the following running parameters are preferred: `blastall -p blastp -d swissprotodb -e 10 -G 1 -E 1 -v 50 -b 50 -I queryseq -o results`

`-p` Program Name [String]
`-d` Database [String]

- e Expectation value (E) [Real]
- G Cost to open a gap (zero invokes default behaviour) [Integer]
- E Cost to extend a gap (zero invokes default behaviour) [Integer]
- v Number of one-line descriptions (v) [Integer]
- 5 -b Number of alignments to show (b) [Integer]
- i Query File [File In]
- o BLAST report Output File [File Out] Optional

The “hits” to one or more database sequences by a queried sequence produced by BLASTN, BLASTP, FASTA, 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 and FASTA algorithms also produce “Expect” or E values for alignments. The E value indicates the number of hits one can “expect” to see over a certain number of contiguous sequences by chance when searching a database of a certain size. The Expect value is used as a significance threshold for determining whether the hit to a database, such as the preferred EMBL database, indicates true similarity. For example, an E value of 0.1 assigned to a hit is interpreted as meaning that in a database of the size of the EMBL database, one might expect to see 0.1 matches over the aligned portion of the sequence with a similar score simply by chance. By this criterion, the aligned and matched portions of the sequences then have a 90% probability of being the same. 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 the EMBL database is 1% or less using the BLASTN or FASTA algorithm.

According to one embodiment, “variant” polynucleotides, with reference to each of the polynucleotides of the present invention, preferably comprise sequences having the same number or fewer nucleic acids than each of the polynucleotides of the present invention and producing an E value of 0.01 or less when compared to the polynucleotide of the present invention. That is, a variant polynucleotide is any sequence that has at least a 99% probability of being the same as the polynucleotide of the present invention, measured as having an E value of 0.01 or less using the BLASTN or FASTA algorithms set at the parameters discussed above.

Variant polynucleotide sequences will generally hybridize to the recited polynucleotide sequence under stringent conditions. As used herein, "stringent conditions" refers to 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
5 washes of 30 minutes each in 0.2X SSC, 0.1% SDS at 65°C. The variant polynucleotide sequences of the invention are at least 50 nucleotides in length.

Variant polynucleotides also include sequences which encode a polypeptide that has a sequence identity of at least 60%, generally 70%, preferably 80%, more preferably 90%,
10 even more preferably 95%, very preferably 98% and most preferably 99% or more to the nucleotide of its respective native nucleotide sequence given in the sequence listing herein.

In general, sequences that code for the α -amylases, starch binding domains/motifs, plastid targeting signals and other polypeptides of the invention will be at least 50%, generally at
15 least 60%, preferably 70%, and even 80%, 85%, 90%, 95%, 98%, most preferably 99% homologous or more with the disclosed sequence. That is, the sequence similarity may range from 50% to 99% or more.

Also encompassed by the invention are fragments of the polynucleotide and polypeptide
20 sequences of the invention. Polynucleotide fragments may encode protein fragments which retain the biological activity of the native protein. Alternatively, fragments used as hybridisation probes generally do not encode biologically active sequences. Fragments of a polynucleotide may range from at least 15, 20, 30, 50, 100, 200, 400 or 1000 contiguous nucleotides up to the full length of the native polynucleotide sequences disclosed herein.

25

Fragments of the polypeptides of the invention will comprise at least 5, 10, 15, 30, 50, 75, 100, 150, 200, 400 or 500 contiguous amino acids, or up to the total number of amino acids in the full length polypeptides of the invention.

30 Variant is also intended to allow for rearrangement, shifting or swapping of one or more nucleotides or domains/motifs (from coding, non-coding or intron regions) from genes (including α -amylases) from the same or other species, where such variants still provide a functionally equivalent protein or polypeptide of the invention or fragment thereof.

It is of course expressly contemplated that homologs to *Mdamy10*, *Mdamy11*, *Atamy3*, and *Osamy10*, exist in other plants. Such homologs are also “functionally equivalent variants” of *Mdamy10*, *Mdamy11*, *Atamy3*, and *Osamy10*, as the phrase is used herein. There are a number of examples of homologs of these genes; several are described in the experimental section. *Mdamy10*, *Atamy3*, and *Osamy10* are all full-length sequences. *Mdamy11* and sequences from Kiwifruit (*Actinidia chinensis*), Blueberry (*Vaccinium corymbosum*), Coffee (*Coffea arabica*), Cotton (*Gossypium hirsutum*), and Rose (*Rosa woodsii*), are all partial sequences. The extent of each sequence is shown graphically in Figure 1, relative to the apple α -amylase *Mdamy8*, which is a cytosol-targeted α -amylase of Family two, which does not have an equivalent to the N-terminal domain of Family three α -amylases. Table 3 shows the percentage identity to *Mdamy10* for each homolog, at both the nucleotide and amino acid level.

Table3:

Homolog	% Nucleotide identity	% Amino acid identity
<i>Mdamy11</i>	91	86
<i>Atamy3</i>	70	68
<i>Osamy10</i>	63	61
Kiwifruit	70	71
Blueberry ^a	81	80
Coffee ^a	77	80
Cotton ^a	79	79
Rose ^a	81	81

^a These partial sequences represent the α -amylase domain of the protein, which has greater amino acid conservation than the N-terminal domain, so that they appear to produce higher amino acid identity scores, and lower nucleotide identity scores, than would be expected for their full-length sequence.

A number of ESTs representing homologs of *Mdamy10*, *Mdamy11*, *Atamy3*, and *Osamy10*, can be found in public databases. Some of these ESTs represent the N-terminal domain of *Mdamy10*, and include part or all of a plastid targeting signal and/or starch binding domain; examples of these ESTs (with Genbank accession numbers listed):

Arabidopsis thaliana: AV528233.1, AV530222.1, AV530101.1

Glycine max: BI698974.1, BE801868.1

Hordeum vulgare: BG299837.1

Lycopersicon hirsutum: AW616947.1, AW616948.1

Lycopersicon pennellii: AW399728.1

Medicago truncatula: BG457945.1, AW689974.1, AW690566.2, AW691339.2, BF639266.1, BF639264.1, BE322876.2, BF642292.1

Solanum tuberosum: BG598624.1

5 *Stevia rebaudiana*: BG523414.1

Other ESTs represent the α -amylase domain of *Mdamy10*; examples of these ESTs (with Genbank accession numbers listed):

Arabidopsis thaliana: AA042154.1, AV524237.1

10 *Glycine max*: AI938911.1, BE800170.1, BM187795.1, BM187817

Hordeum vulgare: AV933515.1, AV912828.1

Lycopersicon esculentum: AW223546.1, BE436615.1, BF176617.1, BE460819.1

Medicago truncatula: BF640476.1

Mesembryanthemum crystallinum: BE035942.1

15 *Pinus taeda*: AI725250.1

Solanum tuberosum: BE922066.1, BG598624.1

Stevia rebaudiana: BG525949.1

Zea mays: BG842601.2, BG840332.1

20 This EST list contains sequences from dicotyledonous and monocotyledonous angiosperms, and also from the gymnosperm *Pinus taeda*. It is expressly contemplated that homologs of *Mdamy10*, *Mdamy11*, *Atamy3*, and *Osamy10*, will exist, other than those listed.

A polynucleotide sequence of the invention may further comprise one or more additional
25 sequences encoding one or more additional polypeptides, or fragments thereof, so as to encode a fusion protein. Chimeric genetic constructs including sequences encoding the novel N-terminal region of a polypeptide of the invention, or one or more starch binding domains/motifs, can be used to produce chimeric protein which can bind to starch. Systems for such recombinant expression include, but are not limited to, mammalian, bacterial, fungal
30 and insect systems.

DNA sequences from plants other than *Malus x domestica* which are homologs of *Mdamy10* and *Mdamy11* may be identified (by computer-aided database searching) and isolated following high throughput sequencing of cDNA libraries prepared from such plants.

35 Alternatively, oligonucleotide probes based on the sequences for *Mdamy10* and *Mdamy11*

provided in SEQUENCE ID No's 1, 9, 11 and 13, can be synthesized and used to identify positive clones in either cDNA or genomic DNA libraries from other plants by means of hybridization or PCR techniques. Probes should be at least about 10, preferably at least about 15 and most preferably at least about 20 nucleotides in length. Hybridization and PCR techniques suitable for use with such oligonucleotide probes are well known in the art. Positive clones may be analyzed by restriction enzyme digestion, DNA sequencing or the like.

The polynucleotides of the present invention may be generated by synthetic means using techniques well known in the art. Equipment for automated synthesis of oligonucleotides is commercially available from suppliers such as Perkin Elmer/Applied Biosystems Division (Foster City, CA) and may be operated according to the manufacturer's instructions.

The primary importance of identification of the polypeptide/polynucleotides of the invention is that they enable modulation of starch content in plants, and particularly plant plastids such as chloroplasts. Modulation may involve a reduction in the expression and/or activity (i.e. silencing) of the polypeptide.

Any conventional technique for effecting this can be employed. Intervention can occur post-transcriptionally or pre-transcriptionally. Further, intervention can be focused upon the gene itself or on regulatory elements associated with the gene and which have an effect on expression of the encoded polypeptide. "Regulatory elements" is used here in the widest possible sense and includes other polynucleotides/polypeptides which interact with the polynucleotides/polypeptides of interest.

Pre-transcription intervention can involve mutation of the gene itself or of its regulatory elements. Such mutations can be point mutations, frameshift mutations, insertion mutations or deletion mutations. These latter mutations include so called "knock-out" mutations in which expression of the gene is entirely ablated.

Examples of post-transcription interventions include co-suppression or anti-sense strategies, a dominant negative approach, or techniques which involve ribozymes to digest, or otherwise be lethal to, RNA post-transcription of the target gene.

Co-suppression can be effected in a manner similar to that discussed, for example, by Napoli *et al.* (1990) and de Carvalho Niebel *et al.* (1995). In some cases, it can involve overexpression of the gene of interest through use of a constitutive promoter. It can also involve transformation of a plant with a non-coding region of the gene, such as an intron from the gene or 5' or 3' untranslated region (UTR).

Anti-sense strategies involve expression or transcription of an expression/transcription product being capable of interfering with translation of mRNA transcribed from the target gene. This will normally be through the expression/transcription product hybridising to and forming a duplex with the target mRNA.

The expression/transcription product can be a relatively small molecule and still be capable of disrupting mRNA translation. However, the same result is achieved by expressing the whole polynucleotide in an anti-sense orientation such that the RNA produced by transcription of the anti-sense oriented gene is complementary to all or part of the endogenous target mRNA.

Anti-sense strategies are described generally by Robinson-Benion *et al.* (1995) and Kawasaki *et al.* (1996).

Genetic constructs designed for gene silencing may include an inverted repeat. 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'

The transcript formed may undergo complementary base pairing to form a hairpin structure provided there is a spacer of at least 3-5 bp between the repeated regions.

Another approach is to develop a small antisense RNA targeted to the transcript equivalent to an miRNA (Llave *et al.*, 2002) that could be used to target gene silencing.

30

Dominant negative approaches involve the expression of a modified plastid α -amylase polypeptide which includes a starch binding domain but lacks a catalytic domain. The result is that the protein binds starch as intended but fails to digest it, while at the same time blocking the binding of the endogenous α -amylase.

The ribozyme approach to regulation of polypeptide expression involves inserting appropriate sequences or subsequences (eg. DNA or RNA) in ribozyme genetic constructs (McIntyre 1996). Ribozymes are synthetic RNA molecules that comprise a hybridizing region
5 complementary to two regions, each of which comprises at least 5 contiguous nucleotides of a mRNA molecule encoded by one of the inventive polynucleotides. Ribozymes possess highly specific endonuclease activity, which autocatalytically cleaves the mRNA.

Alternately, modulation may involve an increase in the expression and or activity of the
10 polypeptide by over-expression of the polynucleotide, or by increasing the number of copies of the polynucleotide in the genome of the host.

To give effect to the above strategies, the invention also provides genetic constructs. The genetic constructs include the intended DNA (such as one or more copies of a polynucleotide
15 sequence of the invention in a sense or anti-sense orientation or a polynucleotide encoding the appropriate ribozyme), a promoter sequence and a termination sequence (which control expression of the gene), operably linked to the DNA sequence to be transcribed. The promoter sequence is generally positioned at the 5' end of the DNA sequence to be transcribed, and is employed to initiate transcription of the DNA sequence. Promoter
20 sequences are generally found in the 5' non-coding region of a gene but they may exist in introns (Luehrsen 1991) or in the coding region.

A variety of promoter sequences which may be usefully employed in the genetic constructs of the present invention are well known in the art. The promoter sequence, and also the termination sequence, may be endogenous to the target plant host or may be exogenous,
25 provided the promoter is functional in the target host. For example, the promoter and termination sequences may be from other plant species, plant viruses, bacterial plasmids and the like. Preferably, promoter and termination sequences are those endogenously associated with the α -amylase genes.

30 Factors influencing the choice of promoter include the desired tissue specificity of the genetic construct, and the timing of transcription and translation. For example, constitutive promoters, such as the 35S Cauliflower Mosaic Virus (CaMV 35S) promoter, will affect the transcription in all parts of the plant. Use of a tissue specific promoter will result in production of the desired sense or antisense RNA only in the tissue of interest. With genetic
35 constructs employing inducible promoter sequences, the rate of RNA polymerase binding and

initiation can be modulated by external stimuli, such as light, heat, anaerobic stress, alteration in nutrient conditions and the like. Temporally regulated promoters can be employed to effect modulation of the rate of RNA polymerase binding and initiation at a specific time during development of a transformed cell. Preferably, the original promoters from the gene in question, or promoters from a specific tissue-targeted gene in the organism to be transformed are used. Other examples of promoters which may be usefully employed in the present invention include, mannopine synthase (mas), octopine synthase (ocs) and those reviewed by Chua *et al.* (1989).

10 The termination sequence, which is located 3' to the DNA sequence to be transcribed, may come from the same gene as the promoter sequence or may be from a different gene. Many termination sequences known in the art may be usefully employed in the present invention, such as the 3' end of the *Agrobacterium tumefaciens* nopaline synthase gene. However, preferred termination sequences are those from the original gene or from the target *Malus* species to be transformed.

The genetic constructs of the present invention may also contain a selection marker that is effective in cells, to allow for the detection of transformed cells containing the genetic construct. Such markers, which are well known in the art, typically confer resistance to one or more toxins. One example of such a marker is the NPTII gene whose expression results in resistance to kanamycin or hygromycin, antibiotics which are usually toxic to plant cells at a moderate concentration (Rogers *et al.* 1988). Alternatively, the presence of the desired genetic construct in transformed cells can be determined by means of other techniques well known in the art, such as PCR or Southern blotting.

25 Techniques for operatively linking the components of the inventive genetic constructs are well known in the art and include the use of synthetic linkers containing one or more restriction endonuclease sites as described, for example, by Maniatis *et al.* (1989). The genetic construct may be linked to a vector capable of replication in at least one system, for example, *E. coli*, whereby after each manipulation, the resulting genetic construct can be sequenced and the correctness of the manipulation determined.

The genetic constructs of the present invention may be used to transform a variety of plants including agricultural, ornamental and horticultural plants. In a preferred embodiment, the

genetic constructs are employed to transform apple, banana, kiwifruit, pine, tomato, cotton, rose, olive, rice, blueberry, *Arabidopsis*, and potato plants.

As discussed above, transformation of a plant with a genetic construct including an open reading frame comprising a polynucleotide sequence of the invention wherein the open reading frame is orientated in a sense direction can, in some cases, lead to a decrease in expression of the polypeptide by co-suppression. Transformation of the plant with a genetic construct comprising an open reading frame or a non-coding (untranslated) region of a gene in an anti-sense orientation will lead to a decrease in the expression of the polypeptide in the transformed plant.

It will also be appreciated that transformation of other non-plant hosts is feasible, including well known prokaryotic and eukaryotic cells such as bacteria (e.g. *E. coli*, *Agrobacterium*), fungi, insect, and animal cells is anticipated. This would enable production of recombinant polypeptides of the invention or variants thereof. The use of cell-free systems (e.g. Roche Rapid Translation System) for production of recombinant proteins is also anticipated (Zubay, 1973).

The polypeptides and proteins of the invention produced in any such hosts may be isolated and purified from same using well known techniques.

Techniques for stably incorporating genetic constructs into the genome of target plants are well known in the art and include *Agrobacterium tumefaciens* mediated introduction, electroporation, protoplast fusion, injection into reproductive organs, injection into immature embryos, high velocity projectile introduction, floral dipping and the like. The choice of technique will depend upon the target plant to be transformed.

Once the cells are transformed, cells having the genetic construct incorporated into their genome may be selected by means of a marker, such as the kanamycin resistance marker discussed above. Transgenic cells may then be cultured in an appropriate medium to regenerate whole plants, using techniques well known in the art. In the case of protoplasts, the cell wall is allowed to reform under appropriate osmotic conditions. In the case of seeds or embryos, an appropriate germination or callus initiation medium is employed. For explants, an appropriate regeneration medium is used.

In addition to methods described above, several methods are known in the art for transferring genetic constructs into a wide variety of plant species, including gymnosperms angiosperms, monocots and dicots (*see, e.g.,* Glick and Thompson (1993) Birch (1997) and Forester *et al.* (1997)). For a review of regeneration of trees, see Dunstan *et al.* (1995).

5

The resulting transformed plants may be reproduced sexually or asexually, using methods well known in the art, to give successive generations of transgenic plants.

10

The nucleotide sequence information provided herein will also be useful in programs for identifying nucleic acid variants from, for example, other organisms or tissues, particularly plants, and for pre-selecting plants with mutations in *Mdamy10* or *Mdamy11* or their equivalents which renders those plants useful in an accelerated breeding program to produce plants in which starch content is modulated. More particularly, the nucleotide sequence information provided herein may be used to design probes and primers for probing or amplification of *Mdamy10*, *Mdamy11* or variants thereof. An oligonucleotide for use in probing or PCR may be about 30 or fewer nucleotides in length. Generally, specific primers are upwards of 14 nucleotides in length. For optimum specificity and cost effectiveness, primers of 16-24 nucleotides in length are preferred. Those skilled in the art are well versed in the design of primers for use in processes such as PCR.

20

If required, probing can be done with entire restriction fragments of the gene disclosed herein. Naturally, sequences based upon Figure 2 or the complements thereof can be used.

Such probes and primers also form aspects of the present invention.

25

Methods to find variants of the of polynucleotides of the invention from any species, using the sequence information provided by the invention, include but are not limited to, screening of cDNA libraries, RT-PCR, screening of genomic libraries and computer aided searching of EST, cDNA and genomic databases. Such methods are well known to those skilled in the art.

30

The invention will now be illustrated with reference to the following non-limiting experiments.

EXPERIMENTAL

35

Materials and Methods**Oligonucleotides used in this work:**

Sequences are from 5' to 3' end:

5	SP1	AACGGTGATTCAGAACAGCATC
	SP2	ATGTACCCTTGAGGCGACACAG
	SP3	GGTGGGAACCAAATCACAGTG
	Oligo-dT	GACTCGAGTCGACATCGATTTTTTTTTTTTTTTTTT
	Race1	GACTCGAGTCGACATCG
10	Atamy1exon3F	GTTACTTACCGGGAAAGCTATAC
	Atamy1exon3R	GGGCGCTCCATCAAAATC
	Atamy2exon5F	TCTTCCACAGGACCTTTATTC
	Atamy2exon5R	AGTCCTCCGGTACAAGAAGTC
	Atamy3exon8F	GGAACAATTGATGAGCTAAAAGATAC
15	Atamy3exon8R	GGAAATGAGGGTCATCTGC
	Amyten12-Xh	TCTCGAGCGTACCATGTCGACGGTTAG
	Amyten13r-Kp	AAGGTACCCATCTTTCCCTCCACCACTTC
	Amyten14r-Kp	AAGGTACCTCTGTCACGTAGCTGAGCATC
	NewUNIf2	ATATTRTGCCAAGGTTTTAACTGGGA
20	NewUNIr2	WARRCGWCCWCCAAAWAKATTCCA

Sampling of plant material

Mature apple fruit (*Malus x domestica*) were harvested from mature apple trees in Auckland,
 25 New Zealand. These fruit were placed in standard commercial boxes and stored at 4 °C for
 eight days. Tissue samples were taken and frozen immediately in liquid nitrogen and stored
 at -80 °C until being used for extraction of RNA.

RNA extractions

30 Between 2g (leaf, flower, young fruit) and 10g (mature fruit) of tissue were ground under
 liquid nitrogen, added to 15mL of heated lysis buffer, and RNA extracted according to the
 method of Langenkamper et al. (1998). cDNA was prepared from mRNA in a standard
 manner.

EST genetic construction and sequencing

cDNAs were cloned into standard library vectors eg Lambda Zap 2 and Lambda Zap Express (Stratagene). Individual clones were excised and plasmid DNA prepared using standard methods. Sequence was then determined by sequencing from both the 5' and 3' ends in the standard manner and intermediate sequences were determined using PCR primers designed to the known sequence in a standard manner.

5' RACE (rapid amplification of cDNA ends)

5' RACE was performed on RNA from apple floral buds, using three *Mdamy10* specific primers, SP1, SP2, and SP3, located at what was the 5' end of the *Mdamy10* sequence at the time. The three primers are reverse-complements of the *Mdamy10* coding sequence, and are nested; SP1 is the most 3' sequence, SP3 the most 5'.

SP1 was used to perform the initial reverse transcription step, performed by Superscript II Reverse Transcriptase (Invitrogen) at 42 °C for 30 min. The product was digested with RNase H and then purified with a PCR purification kit (Qiagen), before having a poly-A tail added with terminal transferase and dATP. The first PCR round used the SP2 and RACE oligo-dT primers, and was performed with Expand Hi-Fidelity enzyme (Roche). The resulting mix was again cleaned up with the Qiagen PCR purification kit, and then used as template for the final PCR round, with SP3 and RACE1 primers. The products of 5' RACE were cloned into pGEM-T Easy (Promega) and sent for sequencing.

RT-PCR of *Arabidopsis* α -amylase genes:

Arabidopsis RNA samples were treated with DNase I (Roche) and then amplified using the Titan One Tube RT-PCR System (Roche), using the primer pairs *Atamy1*exon3(F/R), *Atamy2*exon5(F/R), and *Atamy3*exon8(F/R). These primer pairs amplify the *Arabidopsis* genes *Atamy1*, *Atamy2*, and *Atamy3*, respectively, within domain B. The reaction mix was incubated at 42 °C for 30 min, before undergoing 30 rounds of standard PCR cycling conditions.

Amplification of Family three gene sequences from other plant species

Genomic DNA was extracted from a number of plant tissues, including: needles of pine tree (*Pinus radiata*); sprouting potato tubers (*Solanum tuberosum*); sprouting onion bulbs (*Allium cepa*); rose petals (*Rosa woodsii*); olive leaves (*Olea europaea*); coffee leaves (*Coffea arabica*); banana leaves (*Musa acuminata*); and mango fruit (*Mangifera indica*). Extraction

was performed with a Dneasy Plant Mini Kit (Qiagen), according to the manufacturers instructions. DNA from cotton (*Gossypium hirsutum*) was kindly supplied by Dr John Lunn (CSIRO Plant Industry, Canberra, Australia).

- 5 PCR was performed with Expand Hi-Fidelity enzyme (Roche), using primers NewUNIf2 and NewUNIr2. These degenerate primers were designed from an alignment of *Atamy3*, *Mdamy10*, and *Osamy10*, to amplify only Family three genes, but from a wide range of species. The products of PCR were cloned into pGEM®-T Easy (Promega) and sequenced.

10 **Computational analysis of sequence information:**

Computational analysis was performed using the Wisconsin Package Version 10.1, (Genetics Computer Group (GCG), Madison, Wisc.). Sequence identity was calculated using the pairwise alignment program Gap, which uses the algorithm of Needleman and Wunsch (J. Mol. Biol. 48; 443-453 (1970)). The default parameters were used: Gap creation penalty: 50
15 (for nucleotide input); or 8 (for amino acid input); Gap extension penalty: 3 (nucleotide); 2 (amino acid). Amino acid sequence alignments were performed using the program CLUSTALW (Thompson et al., 1994), and trimmed and shaded using the program GeneDoc (Nicholas and Nicholas, 1997).

20 **Prediction of subcellular localisation of Family three α -amylases**

The subcellular localisations of *Mdamy10*, *Mdamy11*, *Atamy3*, and *Osamy10*, were predicted using servers at the Center for Biological Sequence Analysis (<http://www.cbs.dtu.dk/services/>). The program Target P (Emanuelsson et al., 2000) was used initially to predict the targeting of the proteins. The program ChloroP (Emanuelsson et al. 1999) was then used to predict the length of plastid targeting signal for each protein.
25

Genetic construction of plasmids for expression in plants and *Escherichia coli*

Mdamy10 sequences were cloned or amplified from HortResearch EST genetic construct '62629'. GFP fusion genetic constructs were produced by cloning PCR products into the
30 plasmid pDS-GFP-ART7, which has the gene encoding smGFP (Davis and Vierstra, 1998), driven by the CaMV 35S promoter. The smGFP gene was cloned into pDS-GFP-ART7 so as to leave a *XhoI* and a *KpnI* restriction endonuclease site upstream of the smGFP gene, with the *KpnI* site in-frame with the ATG start of the gene. *E. coli* expression vectors were produced in the vectors pET-30a and pET-30b (Novagen).

PCR was used to amplify the fragments 'cTP' and 'SBD' for fusion to smGFP. The cTP (plastid targeting peptide) product was amplified using primers Amyten12-Xh, located at the 5' end of the coding sequence, containing a *Xho*I restriction endonuclease site upstream of the ATG start of *Mdamy10*, to facilitate cloning; and Amyten13r-Kp, located approximately 300bp downstream of the ATG start site, containing a *Kpn*I restriction endonuclease site, in frame with open-reading frame of *Mdamy10*. This amplified region encodes the plastid targeting signal of *Mdamy10*, plus another 35 amino acids following the signal.

The 'SBD' (starch binding domain) product was amplified using primers Amyten12-Xh, the same primer as for the cTP product; and Amyten14r-Kp, located approximately 1250bp downstream of the ATG start site, containing a *Kpn*I restriction endonuclease site, in frame with open-reading frame of *Mdamy10*; this primer also removes a naturally occurring *Kpn*I site found in the *Mdamy10* sequence. This amplified region encodes the putative cTP, plus both copies of the putative starch binding domain motif pair.

Following PCR, both products were cloned into the vector pGEM®-T Easy (Promega), and excised by digestion with *Xho*I (New England Biolabs) and *Kpn*I (Roche) restriction endonucleases. The restriction products were then cloned into pDS-GFP-ART7, which had been similarly digested with *Xho*I and *Kpn*I. The resulting plasmids were termed pcTP-ART7 and pSBD-ART7. The expression regions of these two genetic constructs were excised by digestion with *Not*I and subsequent cloning into the vector pART27, which had also been cut with *Not*I. This produced the final expression vectors, pcTP-ART27 and pSBD-ART27.

Mdamy10 was removed from EST genetic construct 62629 by digestion with restriction endonucleases *Eco*RI and *Xho*I. This removed the first 220bp of the *Mdamy10* coding sequence, including the cTP encoding region of the cDNA and a further 13 amino acids. This fragment was cloned into pET-30a, which had been cut using the same enzymes, creating the genetic construct pAmy10ET-30a. This added a short peptide to the N-terminus of *Mdamy10*, which includes a His₆ tag, allowing for nickel ion affinity purification.

SmGFP was cloned from pDS-GFP-ART7 into pET-30b using *Kpn*I and *Hind*III restriction endonuclease sites. An *Mdamy8*-GFP fusion (Nterm-GFP) was also cloned into pET-30b in this manner. SBD-GFP was cloned into pET-30a from pSBD-ART27 using *Eco*RI and *Hind*III restriction endonuclease sites, thereby removing the cTP domain from the genetic

construct. Again, all cloning strategies added the His₆ tag to the N-terminus of the expressed proteins.

Preparation of competent *Agrobacterium tumefaciens* GV3101:

- 5 A single colony of *Agrobacterium*, which had been grown on an LB plate containing 25µg/mL of gentamycin, was used to inoculate 100mL of liquid LB media, also containing 25µg/mL of gentamycin. The culture was incubated for 48hr with shaking at 200 rpm in a 28 °C incubator. The cells were harvested by centrifugation (3800 x g for 5 min at 4 °C) and resuspended in 50mL of ice-cold 10% glycerol. The cells were washed twice more in
10 glycerol, before finally being resuspended in 1mL of ice-cold 10% glycerol and aliquoted into 45µL in microcentrifuge tubes. The aliquots were stored at -80 °C.

Transformation of *Agrobacterium tumefaciens* GV3101

- 45 µL aliquots of competent *Agrobacterium* cells were thawed gently on ice. 50-200 ng of
15 plasmid DNA was added to each aliquot and gently mixed, then 40 µL of the cell/plasmid mixture was pipetted into a pre-chilled electroporation cuvette (0.2 cm gap, Bio-Rad). The cells were electroporated using a BioRad GenePulser, on the following settings:

Voltage: 2.5 kV

Capacitance: 25 µF

- 20 Resistance: 400 Ohms

The time constant for the pulse was typically 7-9 ms.

The cells were immediately recovered by addition of 1 mL LB media, then decanted into sterile 15 mL centrifuge tubes and incubated at room temperature, with shaking (60 rpm).

- 25 After 2 hours, 10 µL and 100 µL of the transformed bacteria was spread onto separate LB plates containing rifampicin (10 µg/mL); gentamycin (25 µg/mL); and spectinomycin (100 µg/mL); then grown for 48 hours at 28-30°C.

Transient transformation of *Nicotiana benthamiana* leaves

- 30 Transformation was performed by the method of Hawes et al. (2000). *Agrobacterium tumefaciens* were grown in 5mL cultures of LB media containing rifampicin (10 µg/mL); gentamycin (25 µg/mL); and spectinomycin (100 µg/mL) for 48 hours. The cells were collected by centrifugation and resuspended, to a final OD₆₀₀ of 0.5-0.6, in infiltration medium: 50 mM MES pH5.6, 0.5% (w/v) glucose, 2 mM Na₃PO₄, 100 µM acetosyringone

(added fresh from 200 mM stock in DMSO). The bacterial suspension was injected through the stomata on the underside of *N. benthamiana* leaves, using a “blunt” 1mL syringe (i.e. with no needle attached). The leaves were detached and kept in dark, moist conditions at room temperature for 2-3 days.

5

Microscopy

Leaf samples were examined using an Olympus Vanox AHB3 microscope. Fluorescence was detected using one of two filter sets:

UV filter set - Olympus code BH2-DMU: excitation 320-380 nm, emission 420 nm.

10 GFP plant - Omega filter set XF100-2: excitation 455-495nm, emission 515-560 nm.

Expression of protein in *E. coli*

All pET-30 based expression vectors were transformed following the manufacturer's instructions into BL21 CodonPlus RIL cells (Stratagene) for expression. A single colony was
 15 picked from a plate and used to inoculate 15mL of 2YT media (16 g/L bactotryptone, 10 g/L yeast extract, 5 g/L NaCl, pH 7), containing 30 µg/mL kanamycin and 50 µg/mL chloramphenicol. Cultures were grown overnight at 37 °C, with vigorous shaking (150rpm). Each 5 mL of the 15mL culture was used to seed one of three 330 mL aliquots of 2YT (including 30 µg/mL kanamycin and 50 µg/mL chloramphenicol), in 1 L flasks. These were
 20 shaken at 37 °C for 3-4 hours, until the OD₆₀₀ of the cultures had reached ~1.0, then the flasks and incubator were cooled to 18 °C and expression was induced by addition of IPTG to a final concentration of 0.5 mM. Cells were grown at 18°C with continual shaking, for a further 24 hours before harvesting.

25 Extraction of soluble protein from *E. coli* cells.

Cells were harvested by centrifugation (2500 x g for 10 min); cells from 1 L of culture were then resuspended in 20 mL of the appropriate buffer. For α-amylase activity gels, the cells were resuspended in activity buffer (100 mM Hepes pH 7.5, 5 mM calcium acetate, 10 mM DTT); GFP genetic constructs were resuspended in binding buffer (5mM imidazole, 0.5 M
 30 NaCl, 20 mM Tris-HCl pH 7.9). All buffers contained one Complete protease inhibitor tablet, EDTA free (Roche) per 50 mL of buffer.

The bacterial cells were lysed by twice passing the 20 mL of cell suspension through a French Pressure Cell Press (American Instrument Co. Inc, Silver Spring, Maryland, USA), at

pressure of 12,700 psi. The lysate was centrifuged at 12000 x g for 20 min, and then passed through a filter 0.45 µm pore size. The resulting supernatant was termed the crude extract.

Partial purification of recombinant protein from crude extract

5 For the protein extractions performed in α-amylase buffer, approximately 10 mL of crude extract was applied to a PD-10 gel filtration column that had been pre-equilibrated with binding buffer. GFP samples were not subject to buffer exchange. Purification was performed on a 5 mL Hi-Trap Chelating HP column (Amersham-Pharmacia Biotech) that had been charged with NiSO₄. 10 mL of the extract was applied to the column, and washed
10 according to the manufacturer's instructions. Recombinant protein was eluted into 2.5 mL fractions with elution buffer (1 M imidazole, 0.5 M NaCl, 20 mM Tris-HCl pH 7.9). α-Amylase and pET-30a were immediately exchanged into activity buffer using a PD-10 column.

15 Denaturing gel electrophoresis and western analysis of protein samples

Crude extract and certain fractions from the protein purification steps, were analysed by standard procedures on a SDS-PAGE gel, consisting of a 10% polyacrylamide separating gel and a 4% polyacrylamide stacking gel. Following electrophoresis, gels were either stained with colloidal coomassie protein stain or transferred to Immobilon-P PVDF membrane
20 (Millipore) by semi-dry electrophoresis. For detection of the His₆ motif, the blots were incubated with anti-His₆ monoclonal antibody (Roche), followed by anti-Mouse IgG-alkaline phosphatase (Stressgen) secondary antibodies. For detection of the GFP motif, blots were incubated with anti-GFP polyclonal antibody, IgG fraction (Molecular Probes), followed by anti-Rabbit IgG-secondary antibodies. Both types of blot were detected using 1-STEP™
25 NBT/BCIP alkaline phosphatase detection reagent (Bio-Lab laboratories).

Protein extraction from apple and Arabidopsis

Leaf samples were removed from seedling apples and full rosette plants of Arabidopsis at the end of a 12 hour light cycle. Leaves were frozen in liquid nitrogen and ground using a mortar
30 and pestle, then were added to extraction buffer (50 mM Hepes pH 7.8, 5 mM calcium acetate, 5mM magnesium chloride, 10 mM DTT, 0.5% (w/v) PVPP). The samples were then filtered through 2 layers of Miracloth (Calbiochem, La Jolla, CA., USA) and centrifuged at 14,000 x g for 10 mins. The supernatant was immediately loaded onto native PAGE gels.

Native gel electrophoresis

Crude extract and certain fractions from the protein purification steps, were analysed on two types of native PAGE gels. Starch gels that contained 2% amylopectin in the separating gel. Samples were electrophoresed for 16-20 hrs at a constant current of 10mA, and following
5 electrophoresis, the gels were washed in distilled water and incubated for 24 hrs in activity buffer 2 (50 mM Hepes pH 7.8, 5 mM calcium acetate, 5mM magnesium chloride, 10 mM DTT, 20 (g/mL cycloheximide).. Starch hydrolysis was detected by staining with iodine solution (180 mM potassium iodide, 50 mM iodine), followed by destaining with fixative (30% methanol, 5% acetic acid). Starch gels were stored at 4 °C, covered by Saran plastic
10 wrap.

Samples were also run on native PAGE gels without amylopectin, for 3-4 hours at a constant current of 20 mA. Protein was then transferred into a separating gel containing 2% amylopectin, by capillary blotting with activity buffer 2. After overnight transfer, the
15 amylopectin containing gel (termed the starch transfer gel) was washed and incubated for a further 24 hours in activity buffer 2. Staining, de-staining and storage was the same as for starch gels.

Results

20

Identification of the *Mdamy10* gene from Apple:

To identify α -amylase genes from Apple, an EST database, containing sequences derived from a number of different apple tissues, was compared to the coding sequence of the known apple α -amylase gene, *Mdamy8* (Wegrzyn et al., 2000). The vast majority of α -amylase-like
25 sequences discovered within the HortResearch database corresponded to a previously uncharacterised α -amylase gene (first identified 17 July 2000). Transcripts of this gene are present in numerous tissue libraries, including floral bud, petal, young fruit and mature fruit libraries. Blast searches against the Genbank database revealed a match to a putative α -amylase from *Arabidopsis*, as well as to a number of ESTs from a variety of plants, including
30 *Arabidopsis*, *Lycopersicon esculentum* (tomato), *Medicago truncatula*, *Glycine max* (soybean), *Solanum tuberosum* (potato) and *Pinus taeda* (loblolly pine). The full-length cDNA sequence of this α -amylase, named *Mdamy10* by the applicants, was obtained by sequencing of a single, full-length EST clone, EST 62629, along with the identification of overlapping sequences from partial ESTs. The coding region of *Mdamy10* is 2706 bp long,
35 encoding 901 amino acids. The last 400 amino acids of *Mdamy10* show modest identity to

both *Mdamy8* (46%) and an α -amylase from *Vigna mungo*, *amyVm1* (48%); however the first 500 amino acids are not found in any characterised α -amylase (Fig. 1).

Identification of a homologue of *Mdamy10* in Apple:

5 The only apple EST sequence that matched the 5' region of *Mdamy10* was from a full length cDNA, displaying 79% nucleotide identity, but only 900 nucleotides in length (Fig. 1). The transcript appears to be the product of a gene homologous to *Mdamy10*, except that the equivalent of the third intron of *Mdamy10* (Table 1) has not been spliced from the transcript. This results in cleavage and polyadenylation of the transcript within the third intron, and
10 potential production of a truncated protein. We have tentatively labelled the transcript *Mdamy11*.

We have two further fragments of *Mdamy11* sequence. The first was produced with a primer pair designed to *Mdamy10* cDNA sequence, which amplified two products of slightly
15 different sizes (1028bp and 831 bp) from genomic DNA. The fragments had very similar exon sequences (95% identity at both nucleotide and amino acid levels), and each contained three introns, found in the same positions. However the introns of the two fragments differ significantly in their sequence (80% identity at nucleotide level) and their size (602bp total, versus 405 bp). The larger fragment (and hence larger intron) is from *Mdamy10*, the smaller
20 is presumed to be from *Mdamy11*. The sequence is from close to the 3' end of the α -amylase domain, and so does not overlap the *Mdamy11* transcript (Fig. 1).

A final additional fragment was isolated during 5' RACE (rapid amplification of cDNA ends) of *Mdamy10* from apple floral bud RNA. The 600 bp region overlapped with sequence from
25 *Mdamy10*, with 6 bp in disagreement within the 95bp of overlap. These were initially thought to be sequence errors, accumulated from both EST genetic construction and 5' RACE, however attempts to continue upstream of this region using 5' RACE were unsuccessful. Finally, sequence from the full-length EST 62629 reached the region of the 5' RACE fragment, and was found to share only 86% identity at the nucleotide level. The 5'
30 RACE fragment was automatically removed from the *Mdamy10* sequence contig. This fragment does not overlap either of the previous *Mdamy11* sequences, but corresponds to the junction of the novel N-terminal domain and the α -amylase domain of *Mdamy10* (Fig. 1).

These three fragments of *Mdamy11*, when taken together, cover 61% of the total coding sequence of *Mdamy10*, with 86% amino acid identity between the two genes through these regions.. The identity is much greater in the C-terminal (α -amylase) domain (96%) than in the N-terminal domain (79%).

5

Identification of a homologue of *Mdamy10* in *Arabidopsis*:

Atamy3 was originally annotated in Genbank by TIGR, predicted from genomic sequence. The annotation suggested a coding sequence 2478 bp in length, encoding a protein of 826 amino acids (compared to around 415 aa for other plant α -amylases). We reanalysed the appropriate chromosomal segment, using Eukaryotic GeneMark.hmm, (LUKASHIN & BORODOVSKY, 1998) to predict gene splicing. This algorithm predicted an even larger coding region of 2661 bp (figure 4), due to inclusion of two extra exons, producing a protein 887 aa in size (figure 5), displaying 68% identity to the predicted *Mdamy10* product (Fig. 1). The N-terminal domain (bold and underlined in figure 5) has 56% identity to *Mdamy10*, while the rest of the protein has 82% identity to *Mdamy10*. The polynucleotide sequences display 70% identity.

Our sequence prediction was confirmed by the sequencing of an *Atamy3* cDNA clone at the Salk Institute, CA., published in Genbank on 21 August 2001 (accession AY050398). *Atamy3* is located on chromosome 1 (BAC T17F3). Blast searches have not yielded any other Family three α -amylase genes in the *Arabidopsis* genome.

Identification of a homologue of *Mdamy10* in rice (*Oryza sativa*):

Osamy10 was originally discovered by using the BlastP program to search the genbank non-redundant database, using *Mdamy10* as the query sequence. The sequence was annotated in Genbank by members of the rice genome research program as a 'putative alpha-amylase'; the coding sequence had been predicted from genomic sequence (accession AP003408 – 23 May 2002). As with *Atamy3*, we reanalysed the appropriate chromosomal segment, using Eukaryotic GeneMark.hmm, to predict gene splicing. This revealed misannotation of the gene, namely the inclusion of a 99 bp DNA segment which was in fact part of an intron (this was confirmed by comparing the relative intron positions of *Atamy3*). The sequence predicted by Eukaryotic GeneMark.hmm is 2619 bp (figure 6), encoding 873 amino acids (figures 1 and 7). The polynucleotide sequence displays 63% identity to *Mdamy10*. The protein sequence displays 61% identity to *Mdamy10*, over its full length, however the N-

terminal domain (bold and underlined in figure 7) has only 44% identity to *Mdamy10*, while the rest of the protein has 80% identity to *Mdamy10*.

Genomic sequences from other plant species:

5 PCR was performed with degenerate primers, to amplify Family three sequences from the genomic DNA of a number of plant species. Figure 10 shows the results of PCR. All DNA samples except onion produced a molecule of about 320 bp, which was the size of product expected from amplification of *Mdamy10* cDNA with the primer pair. Cloning and sequencing of several of these products confirmed them to be *Mdamy10* cDNA sequence,
10 presumably caused by aerosol contamination. The abundance of this product inhibited cloning of other PCR products in the samples, but subsequent experiments failed to pinpoint and/or remove the source of contamination.

Most DNA samples produced at least one other DNA molecule larger than the 320 bp band.
15 At the time of writing, only three of these products have been successfully cloned; an approximately 800 bp product from rose (only partially sequenced) (Fig. 10., lane 4), a 962 bp product from coffee (Fig. 10., lane 6), and an 808 bp product from cotton (Fig. 10., lane 9). All three sequences contain a single intron, which accounts for almost all of the size difference between the products of different species. The position of each intron is conserved
20 between the three products, and correspond to the seventh intron of *Atamy3* (intron 9 of Table 1). This intron position is not found in α -amylases from Families one or two. There were no introns found in conserved positions of Family one and Family two introns.

The three fragments all cover a region of the α -amylase genes that encodes part of the
25 previously characterised domains A and B of α -amylases (Fig. 1). The introns were removed from the sequences and the resulting segments were translated into the expected polypeptides. Both the polynucleotide and polypeptide sequences were aligned with representatives of each of the three α -amylase families; *Mdamy10* (Family three), *Mdamy8* (Family two), and *AmyVm1* (Family one). The degree of identity between each pair was calculated as a
30 percentage (Table 4).

Table 4:

% Identity	<i>Mdamy10</i>		<i>Mdamy8</i>		<i>AmyVm1</i>	
	DNA	AA	DNA	AA	DNA	AA

Coffee	77	80	51	42	56	56
Cotton	79	79	51	40	52	48
Rose	81	81	47	21	54	44

All three fragments show significantly higher identity towards *Mdamy10* than they do towards α -amylases from other families; particularly at the amino acid level.

5 Structure of α -amylase and α -amylase-like genes in plants:

The gene structures of plant α -amylases have been compared previously, most significantly in HUANG et al. (1990). Repeating this type of analysis, but using new *Arabidopsis*, apple and rice α -amylase sequences, produces a far more complicated picture of α -amylase intron evolution. Previously sequenced α -amylase genes have at most three introns. Some cereal
10 genes appear to have lost the second intron (HUANG et al., 1990), but the position of each intron is conserved between genes and species.

The intron/exon structure of *Mdamy10* is significantly different from previously sequenced α -amylase genes (Table 1). The gene contains 12 introns, compared to other characterised plant
15 α -amylases, which have 2 or 4 introns. Of the 12 introns that interrupt the coding sequence of *Atamy3*, 6 of them are within the α -amylase-encoding region of the gene (i.e. the 3' half). None of the introns of the Family three α -amylases correspond to introns of the Family one α -amylases.

20 Post-transcriptional processing of α -amylase-like genes:

In contrast to the *Mdamy8* transcripts, *Mdamy10*, *Mdamy11* and *Atamy3* all appear to have short 5' UTRs - between 19 bp and 46 bp. However, *Mdamy10* has a very long 3' UTR, up to 557 bp, although one of the four different polyadenylation sites can produce a 3' UTR as short as 428 bp. The 3' UTR sequence of *Atamy3* is only 220 bp.

25

Subcellular localisation of plant α -amylases:

All of the Family three proteins were identified as plastid-targeted; the program ChloroP (EMANUELSSON et al., 1999) predicted transit peptide lengths of 61 amino acids (*Mdamy10*), 70 amino acids (*Mdamy11*), 55 amino acids (*Atamy3*), and 53 amino acids
30 (*Osamy10*). To test this, the plastid targeting sequence of *Mdamy10* was fused to green

fluorescent protein and transiently expressed in the leaves of *N. benthamiana*. The pattern of GFP localisation was compared to GFP alone.

GFP is localised to the cytosol of epidermal cells (Fig. 11), and also to the nucleus (visible in the very centre of epidermal cells, top left panel, Fig 11), but not to the vacuole, which dominates the cell's volume. By comparison, the cTP-GFP fusion localises to multiple organelles, discoid in shape, within each cell (Fig. 11, top right and lower panels). Each disc has a maximum diameter of about 4µm, as compared to the nucleus, which is about 10µm in diameter. The same organelles display chlorophyll autofluorescence (red) under UV illumination (Fig. 11, top right panel), identifying them as chloroplasts; co-localisation of the GFP and chlorophyll appears as an orange/yellow (only under UV illumination).

A second fusion genetic construct, SBD-GFP, was produced, containing the majority of the N-terminal domain (including the cTP) of *Mdamy10*, also fused to GFP. This too was expressed in *N. benthamiana* leaves, but produced no detectable fluorescence. It was unclear whether this was due to improper folding of the GFP domain or because of degradation of the protein in plant cells. We attempted to determine localisation of the fusion protein by immunological detection using anti-GFP antibodies, but could not detect any significant signal from GFP.

20

Expression of GFP fusion proteins in *E. coli*:

The inserted sequences of pDS-GFP-ART7, pDS-Nterm-ART7 (GFP fused to the N-terminus of *Mdamy8*), and pSBD-ART7 were transferred into pET-30 vectors and expressed in *E. coli*. The cTP of *Mdamy10* was removed from the SBD fusion genetic construct, at a naturally occurring EcoRI site, so as to increase the overall solubility of the protein. After induction with IPTG, samples of the cultures were examined by fluorescence microscopy; green fluorescence was observed for all three proteins, although pGFP-ET-30b produced more intense fluorescence than the other two samples; pET-30a alone produced no fluorescence. This was also true of the soluble protein fractions extracted from the cultures. Extracts were made of the soluble and insoluble fractions of the cell, and the samples were subjected to SDS-PAGE and western analysis as shown in Fig 12.

These analyses show that pGFP-ET-30b produced the most soluble, full-length (31 kDa) protein; this corresponded to the highest level of fluorescence. pNterm-ET-30b produced almost no soluble, full length protein (78 kDa), whilst pSBD-ET-30a had a small amount of

soluble full length product (71 kDa). All fusion proteins appear slightly larger on the gel than their expected sizes. The vast majority of protein produced by pSBD-ET-30a was found in the insoluble fraction. Both of the fusion genetic constructs produced smaller, soluble products, which may be breakdown products of GFP or may come from internal initiation
5 with the fusion gene, leading to translation of an unfused GFP molecule. This GFP may be responsible for the fluorescence seen in induced cultures.

Expression of α -amylase *Mdamy10* in *E. coli*:

Mdamy10 protein was expressed in *E. coli* at low temperatures, and the protein was extracted
10 and partially purified on a nickel column. Both purified and unpurified protein were electrophoresed into polyacrylamide gels containing 2% amylopectin; these gels sort proteins based on their affinity for the amylopectin, proteins with low affinity travel further than high affinity proteins. The gels were stained with iodine, which forms a purple complex with amylopectin; a clear area on the gel indicates hydrolysis of the substrate (Fig. 13).

15 The gels revealed a number of intrinsic *E. coli* proteins with hydrolytic activity, which can be seen in protein extracts from pET-30a expressing bacteria (labelled X, Y and Z in Fig. 13). The same intrinsic activity was visible in bacterial cultures expressing pAmy10-ET-30a, but there were also two extra 2 bands close to the boundary of the stacking and separating gels
20 (labelled 1 and 2 in Fig 13). These bands were visible only in crude protein extracts; desalting or purifying on a nickel column led to loss of this activity.

Crude extract of *Mdamy10* protein was examined on a starch transfer gel, alongside whole protein samples from apple and *Arabidopsis* leaves. In this analysis, proteins are sorted based
25 upon size and charge, rather than their affinity for starch.

Figure 14 shows a starch transfer gel that has been stained with iodine. Extracts from *E. coli* containing pAmy10-ET-30a show a band of amylolytic in the middle of the gel (labelled A in Fig. 14, lanes 2 & 5) that is not present in the pET-30a only control. A faint band of the same
30 size can be seen in extracts of apple leaf tissue (lane 3, Fig 14). The intrinsic *E. coli* activity seen in starch gels is visible only at the very bottom of the starch transfer gels, presumably they have high mobility in the native PAGE gels and have mostly eluted from the gel prior to transfer.

We undertook SDS-PAGE of the various protein fractions, followed by western blotting using anti-His₆ antibodies (Fig. 15). The expected size of the Mdamy10 protein plus the 6x His tag of pET-30a is 99 kDa. A 105 kDa protein, with a 6x His tag was expressed only in pAmy10-ET-30a cultures, and was successfully recovered by nickel column purification. The difference between the size of expected and expressed protein is probably due to the inherent inaccuracy of determining size in this manner. There were also a few secondary products purified, the major protein was around 55 kDa in size. This fragment eluted mostly in the third 2.5 mL elution fraction, while the 105 kDa protein eluted in the second and third fractions. The 55kDa product is probably responsible for the lower band (band 2) in the Mdamy10 lane of starch gels (Fig. 13), as the intensity of band 2 appears to be relative to the amount of 55kDa protein in the extract (data not shown). It appears Mdamy10 was inactivated during purification, rather than being degraded or misplaced.

Expression profile of Mdamy10 and homologues in plants:

ESTs encoding Mdamy10 have been sequenced from a wide variety of apple tissues, including:

Apple fruit skin peel, from tree-ripened fruit 150 days after full bloom (DAFB).

Apple fruit cortex tissue, from tree-ripened fruit 150 DAFB.

Young apple fruit, 10 DAFB.

Young apple fruit, 24 DAFB.

Apple fruit stored at 0.5°C for 24 hours.

Spur buds from apple trees.

Apple pre-opened floral bud.

Apple leaf infected with *Venturia inaequalis*.

ESTs from kiwifruit were found in several tissues from ripe fruit, including skin and inner cortex, and also in breaking bud. A single EST was isolated from the skin of blueberry (*Vaccinium corymbosum*).

Some RT-PCR experiments were performed on a developmental series of tissues from Arabidopsis. The experiment used primers specific to each of the three *Arabidopsis* genes to ascertain the relative expression of each family (Fig. 16).

The RT-PCR shows that *Atamy1* has very low expression levels in *Arabidopsis* (transcript was detectable in some tissues after 40 rounds of PCR (data not shown)), whilst *Atamy2* and

Atamy3 have similar levels of expression, but different patterns of expression. *Atamy3* expression was at its highest in the leaves of growing stems, and moderate in a number of other tissues (emerging cotyledons, whole seedlings, and young seed pods).

5 Discussion:

Sequencing of apple genes and analysis of sequence databases has yielded a small number of atypical α -amylase sequences. Apple has at least six distinct α -amylase-like genes, grouped into three families, each containing two genes, while *Arabidopsis* has only one representative for each family. Apple is a cryptic diploid plant, i.e. it has evolved by the fusion of two ancestral genomes, each with its own set of genes and hence one gene copy from each α -amylase family. Thus we can assume that the three sequence fragments attributed to *Mdamy11* are all from a single gene, as the other Family three gene (*Mdamy10*) has been fully sequenced. We cannot make the same assumptions regarding the sequence fragments from kiwifruit, as we have not sequenced any single family three gene in its entirety.

Perhaps the most striking difference between each gene family is the number of introns within each gene and their relative position. Each of the three families has its own characteristic intron structure, with anywhere from two to twelve introns interrupting the coding sequence; Families one and three do not share any common intron/exon boundaries, although both share boundaries with Family two.

Our investigations have shown that Family three α -amylases are expressed in a variety of different tissues, in a number of plant species. A recent microarray analysis of diurnal and circadian regulated genes in *A. thaliana* (SCHAFFER et al., 2001) identified an EST from *Atamy3* as one of many transcripts displaying a diurnal expression pattern. The *Atamy3* transcript is upregulated in the afternoon, along with a hexose transporter, and repressed again in the early morning. This pattern corresponds to the diurnal breakdown of starch in chloroplasts and subsequent export of sugars from photosynthetic tissues, which takes place at night, and is consistent with the predicted plastid-localisation of *Atamy3*. The expression pattern for *Mdamy10* has not been specifically explored, although its presence in many different tissues has been demonstrated by EST sequencing and some initial microarray experiments. Information from a wide range of plants suggests that Family three α -amylases are expressed in plant tissues involved in degradation of plastid-bound starch, including

photosynthetic cells (during night-time), fruit during maturation and floral buds breaking in spring.

- The 5' and 3' untranslated regions of each gene may be important in post-transcriptional regulation, possibly controlling the stability or localisation of the mRNA transcript. The 3' UTR of a rice α -amylase transcript has been shown to mediate mRNA levels in a sugar-dependent manner, by destabilising the transcript when sugar is abundant (CHAN & YU, 1998). *Mdamy10* has a long 3' UTR that may contain similar mRNA stability elements.
- We were able to amplify sequence of family three genes from several different plant species, using degenerate primers designed from the DNA sequences of *Atamy3*, *Mdamy10*, and *Osamy10*. Sequences from coffee, cotton and rose all showed a high degree of identity to *Mdamy10*, and almost certainly come from family three genes, based on the degree of identity and the position of the introns within each sequence. It is quite likely that the other PCR products shown in figure 10, also represent family three α -amylases, and that the family is ubiquitous in plants. The PCR fragments could be used as probes to isolate full-length α -amylase sequences from the source organisms, as well as related species; they could also be used as molecular markers in plant breeding.
- Mdamy10* was successfully expressed in *E. coli*, and has been shown to possess the ability to enzymatically degrade amylopectin in native gels. Unfortunately the activity of the enzyme appears to be lost upon purifying by nickel ion affinity chromatography. The crude protein extract produces two *Mdamy10* specific bands when electrophoresed into amylopectin containing gels, but only one band when transferred into amylopectin gels from a native PAGE gel. The lower band seen on starch gels could be caused by partially degraded protein; some evidence for this is seen in western blots using anti-His₆ antibodies, where a smaller (55 kDa) protein is eluted slightly later than the 105 kDa *Mdamy10* band. The 55 kDa band may represent the N-terminal half of the *Mdamy10* protein, which would retain the His₆ protein tag, and the C-terminal part of the protein may be released but could still remain enzymatically active. The C-terminal fragment would not be purified by nickel ion affinity chromatography, and would not show up on western blots with anti-His₆ antibody. However, loss of this fragment is not the reason for loss of activity during purification, as activity is lost upon desalting on a PD-10 column (which would not remove the fragment). The C-terminal

fragment would also run much further than the full-length protein, on a native PAGE gel, and may well have eluted from the gel, along with the intrinsic *E. coli* activity seen on starch gels.

5 The lack of a second activity band on starch transfer gels could also be explained if the pool of expressed *Mdamy10* protein did not have a uniform affinity for starch. This could be due to misfolding of one or more of the domains of the protein, or due to some chemical alteration made to the protein before, during, or after extraction from *E. coli*, for example oxidation of amino acid residues critical for binding amylopectin. Alternatively, the *Mdamy10* protein may form multimeric complexes in the presence of amylopectin, reducing its mobility in
10 starch gels and producing multiple activity bands. Starch transfer gels also show that there is an intrinsic amylolytic activity in apple leaves that co-migrates with *Mdamy10* protein expressed in *E. coli*.

The ability of *Mdamy10* to degrade starch, along with its sequence similarity to previously
15 characterised α -amylase genes, largely confirms that it is an active α -amylase.

Expression of an *Mdamy10*-GFP fusion protein in plant cells was able to demonstrate that the N-terminal end of the protein contains a plastid targeting signal that is capable of effecting GFP import into the chloroplasts of *N. benthamiana* cells. This confirms earlier
20 computational predictions of such targeting signals in the protein. The existence of very similar peptide sequences in other Family three proteins indicates that the entire family is plastid targeted. This is significant, as no other α -amylases have been described that are targeted to the plastid, which is the main site of starch storage and degradation within plants.

25 We believe that the Family three α -amylases presently described (including *Mdamy10*, *Mdamy11*, *Atamy3* and *Osamy10*) are the enzymes responsible for degrading all forms of plastid-bound starch, i.e. both diurnal and storage starch.

INDUSTRIAL APPLICATION

30

In its primary aspect, the invention has application in modulating the starch content of organisms including plants, and plant plastids. This family of α -amylases is implicated in the modification and degradation of plastid-bound starch, that is, both diurnal and storage starch. The invention can be used to modify various aspects of organisms including plants. Such

aspects include, starch content, starch composition, starch polymer type, sugar content, ripening, texture, solids content, viscosity of processed tissue, resistance to chilling damage, processing properties, wood quality and yield.

5 Examples of commercially valuable processes where the methods of the invention directed at modulating starch degradation in transgenic plants may be useful include:

- 1) prevention of low temperature sweetening in potato tubers which may require inhibition of starch degradation
- 2) prevention of tuber sprouting which may require inhibition of starch degradation
- 10 3) improvement in the storage of starch-containing fruit such as banana, apple, kiwifruit, papaya and mango, where inhibition of starch degradation may aid storage of the fruit, possibly without a need for low temperature storage.
- 4) dormancy breaking (and germination of seeds) which requires action of starch degrading enzymes to provide energy for new growth.

15

Chimeric genetic constructs according to the invention can be used to target chimeric proteins to plastids and/or starch granules for a variety of purposes including biopharming of vaccines, or targeting of proteins to modify, either plastid or non-plastid, starch in transgenic plants. Such genetic constructs may also be used to target bacterial, fungal or algal amylases to the starch granules and/or plastids of transgenic plants. Opportunities for DNA shuffling also exist using the sequences of this invention.

20

The α -amylases of the invention may of course also be employed in known industrial applications in which starch degradation is required. This includes processing of animal feeds, detergents, food and beverages, textiles, healthcare and brewing.

25

There are also opportunities to produce the α -amylases of the invention, variants of the enzymes or chimeric proteins including the starch binding domain/motifs by fermentation processes, or recombinant expression, for use in industrial applications. For example genetic constructs of the invention could be used to produce chimeric protein including a fungal amylase and starch binding domain/motifs of the invention for more efficient degradation of plant starch in waste processing.

30

WO 03/027141

PCT/NZ02/00200

It will further be appreciated by those persons skilled in the art that the present description is provided by way of example only and that the scope of the invention is not limited thereto.

REFERENCES

- ALTSCHUL, S. F., *et al.*, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* 25:3389-34023 (1997)
- 5 CARVALHO N. *et al* (Plant Cell 7:347-258, 1995).
- CHAN, M.T. & YU, S.M. 1998. The 3' untranslated region of a rice α -amylase gene functions as a sugar-dependent mRNA stability determinant. *Proc. Natl. Acad. Sci. USA* 95: 6543-6547.
- 10 CHUA *et al.* (Science, 244:174-181, 1989).
- DAVIS, S.J. & VIERSTRA, R.D. 1998. Soluble, highly fluorescent variants of green fluorescent protein (GFP) for use in higher plants. *Plant Mol. Biol.* 36: 521-528.
- 15 DUNSTAN *et al.*, Somatic embryogenesis in woody plants. In: Thorpe, T.A. ed. 1995: *in vitro* embryogenesis of plants. Vol 20 in Current Plant Science and Biotechnology in Agriculture, Chapter 12, pp. 471-540.
- 20 EMANUELSSON, O., NIELSEN, H., BRUNAK, S. & VON HEIJNE, G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 300: 1005-1016.
- 25 EMANUELSSON, O., NIELSEN, H. & VON HEIJNE, G. 1999. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Science* 8: 978-984.
- GLICK & THOMPSON, eds., *Methods in Plant Molecular Biology*, CRC Press, Boca Raton, Florida (1993)
- 30 BIRCH, R.G., *Ann Rev Plant Phys Plant Mol Biol.*, 48:297 (1997);
- FORESTER *et al.*, *Exp. Agric.*, 33:15-33 (1997).
- 35

- HAWES, C., BOEVINK, P. & MOORE, I. 2000. Green fluorescent protein in plants. In Protein localization by fluorescence microscopy: a practical approach (ed. V.J. Allan), p. 163-177. Oxford University Press.
- 5 HUANG, N., SUTLIFF, T.D., LITTS, J.C. & RODRIGUEZ, R.L. 1990. Classification and characterization of the rice α -amylase multigene family. *Plant Mol. Biol.* **14**: 655-668.
- LANGENKAMPER, G., MCHALE, R., GARDNER, R.C. & MACRAE, E.A. 1998. Sucrose-phosphate synthase steady-state mRNA increases in ripening kiwifruit. *Plant Mol. Biol.* **36**: 857-869.
- 10 LLAVE, C., XIE, Z., KASSCHAU, K.D. & CARRINGTON, J.C. (2002). Cleavage of *Scarecrow-like* mRNA targets directed by a class of *Arabidopsis* miRNA. *Science*. **297**: 2053-2056.
- 15 LUEHRSEN, K.R., *Mol. Gen. Genet.* **225**:81-93, 1991)
- LUKASHIN, A. & BORODOVSKY, M. 1998. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res.* **26**: 1107-1115.
- 20 MANIATIS *et al.*, (*Molecular Cloning: A Laboratory Manual*, Cold Spring Harbour Laboratories, Cold Spring Harbour, NY, 1989).
- (MCINTYRE CL, MANNERS JM, *Transgenic Res.* **5**(4):257-262, 1996)
- 25 NAPOLI *et al.* (*Plant Cell* **2**:279-290, 1990)
- NEEDLEMAN & WUNSCH (*J. Mol. Biol.* **48**; 443-453 (1970))
- 30 NICHOLAS, K. B. & NICHOLAS, H. B. 1997. GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by the author.
- W. R. PEARSON, "Rapid and Sensitive Sequence Comparison with FASTP and FASTA, "*Methods in Enzymology* **183**:63-98 (1990)
- 35

W. R. PEARSON & D. J. LIPMAN, "Improved Tools for Biological Sequence Analysis",
Proc. Natl. Acad. Sci. USA 85:2444-2448 (1988)

ROBINSON-BENION *et al.*, (1995), Anti-sense techniques, *Methods in Enzymol.*
 5 254(23):363-375 and Kawasaki *et al.*, (1996), *Artific. Organs* 20 (8): 836-848.

ROGERS *et al.*, in *Methods for Plant Molecular Biology*, A Weissbach and H Weissbach eds,
 Academic Press Inc., San Diego, CA (1988)).

10 SCHAFFER, R., LANDGRAF, J., ACCERBI, M., SIMON, V., LARSON, M. & WISMAN,
 E. 2001. Microarray analysis of diurnal and circadian-regulated genes in *Arabidopsis*. *Plant*
Cell 13: 113-123.

THOMPSON, J. D., HIGGINS, D. G. & GIBSON, T. J. 1994. CLUSTALW: improving the
 15 sensitivity of progressive multiple sequence alignment through sequence weighting, position-
 specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673-4680.

WEGRZYN, T., REILLY, K., CIPRIANI, G., MURPHY, P., NEWCOMB, R., GARDNER,
 R. & MACRAE, E. 2000. A novel α -amylase gene is transiently upregulated during low
 20 temperature exposure in apple fruit. *Eur. J. Biochem.* 267: 1313-1322.

ZUBAY, G. (1973). In vitro synthesis of protein in microbial systems. *Annu. Rev. Genet.*
 7, 267-287.

25

GUIDE TO SEQUENCE LISTING

	1	<i>Mdamy10</i> polynucleotide
	2	<i>Mdamy10</i> polypeptide
5	3	<i>Mdamy10</i> polynucleotide - 5' end
	4	<i>Mdamy10</i> polypeptide - N-terminal end (corresponds to 3)
	5	<i>Mdamy10</i> polypeptide - starch binding region (no cTP)
	6	<i>Mdamy10</i> polynucleotide - cTP encoding
	7	<i>Mdamy10</i> polypeptide - cTP
10	8	<i>Atamy3</i> polypeptide - cTP
	9	<i>Mdamy11</i> polynucleotide 5' piece
	10	<i>Mdamy11</i> polypeptide 5'piece
	11	<i>Mdamy11</i> polynucleotide - central piece
	12	<i>Mdamy11</i> polypeptide - central piece
15	13	<i>Mdamy11</i> polynucleotide 3' piece
	14	<i>Mdamy11</i> polypeptide 3'piece
	15	Predicted <i>Atamy3</i> polynucleotide sequence (GeneMark.hmm)
	16	Sequenced <i>Atamy3</i> polynucleotide sequence (Salk institute)
	17	<i>Atamy3</i> polypeptide sequence
20	18	<i>Mdamy10</i> genomic sequence (includes introns)
	19	<i>Mdamy11</i> genomic sequence (includes introns)
	20	Kiwifruit sequence 1 polynucleotide; Equivalent of <i>Mdamy10</i> 5'
	21	Kiwifruit sequence 1 polypeptide; Equivalent of <i>Mdamy10</i> N-terminal end
25	22	Kiwifruit sequence 2 polynucleotide; Equivalent of <i>Mdamy10</i> central region
	23	Kiwifruit sequence 2 polypeptide; Equivalent of <i>Mdamy10</i> central region
	24	Kiwifruit sequence 3 polynucleotide; Equivalent of <i>Mdamy10</i> 3'
30	25	Kiwifruit sequence 3 polypeptide; Equivalent of <i>Mdamy10</i> C-terminal end
	26	Kiwifruit sequence 4 polynucleotide; Equivalent of <i>Mdamy10</i> 3'
	27	Kiwifruit sequence 4 polypeptide; Equivalent of <i>Mdamy10</i> C-terminal end
35	28	Blueberry sequence 1 polynucleotide
	29	Blueberry sequence 1 polypeptide
	30	Coffee polynucleotide sequence
	31	Coffee polypeptide sequence
	32	Cotton polynucleotide sequence
40	33	Cotton polypeptide sequence
	34	Rose polynucleotide sequence
	35	Rose polypeptide sequence
	36	SP1
	37	SP2
45	38	SP3
	39	Oligo-dT
	40	Race1
	41	NewUNIf2
	42	NewUNIr2
50	43	<i>Atamy1</i> exon3F
	44	<i>Atamy1</i> exon3R
	45	<i>Atamy2</i> exon5F
	46	<i>Atamy2</i> exon5R
	47	<i>Atamy3</i> exon8F
55	48	<i>Atamy3</i> exon8R
	49	<i>Amyten12</i> -Xh
	50	<i>Amyten13r</i> -Kp
	51	<i>Amyten14r</i> -Kp
	52	Predicted <i>Osamy10</i> polynucleotide sequence (GeneMark.hmm)
60	53	Predicted <i>Osamy10</i> polypeptide sequence (GeneMark.hmm)
	54	<i>Osamy10</i> genomic sequence, including introns
	55	<i>Atamy3</i> genomic sequence, including introns

SEQUENCE LISTING

<110> THE HORTICULTURE AND FOOD RESEARCH INSTITUTE OF NZ

5 <120> PLANT POLYPEPTIDES AND POLYNUCLEOTIDES ENCODING SAME

<130> P463376 TVG

<140>

10 <141>

<150> NZ 514547

<151> 2001-09-28

15 <160> 55

<170> PatentIn Ver. 2.1

<210> 1

20 <211> 3324

<212> DNA

<213> Malus domestica

<400> 1

25 attccgtcca tcagtttctgt ttcttcagtc gcctgagcgg ccgttgctcg ttccatgtcg 60
acggtttagga tagagcccct cctccaccac taccgtcgac agaaaccag ccaccgcctt 120
ccgccgtcga agcaccatt aaagctcagc tcttctttca ctgcttttcc aaagaagcta 180
gtagtctcca atggccgcag cttctgcaac ttccagcctc ccactctcag tgtcagagct 240
gcctccaccg atacagccac cgtcgaggcc accgaattcg ccgacgcctt ctacaaggag 300
30 accttccctc tcaagcgaac tgaagtgggtg gagggaaaga tgatcgtgaa attagataat 360
gggaaggatg caaagaattg ggtgctgact gtgggttgta atcttcctgg aaaatgggtt 420
cttccactggg gagttaatta tgtggatgac gtccgcagtg aatgggatca gcctcctagt 480
gaaatgagac cagctgggtc agtttccatc aaggactatg caatagagac acccttgaag 540
gaatcgttgt cgccgggtggg aggcgataca tctcacgaag tgaagattga tgttacaccc 600
35 aacagcgcga tcgcagcgat aaattttgtt ctcaaggatg aagaaaccgg tgcctgggat 660
cagcatagag ggagagactt taaagtgcct ttctggtggc acctgcaaga cgatgacaat 720
gtagtgggag caacaagggc cttgggcgcg tggtcaggaa ctttgggaaa actatccaat 780
gtgtttgtca aagcagaaac atcaaattcc aaagatcaag aaagcagcag tgaatccaga 840
gacctcaac agaaaactat gcgtctagaa ggattctatg aagaactgcc aattgcaaaa 900
40 gaaattgctg ttaaccattc agcaactgtt tccgttagga agtgccctga gactactaag 960
aatcttctat acttggaaac agatttacct gatcatgctg ttgttctactg gggagtgttg 1020
agagatgatg ctaaaagatg ggaaattcca gctgccccgc atccaccaga aacagtagtt 1080
ttcaaggaca aggcctttgcg gactcgatta cagcaaaggg aggatggaaa tggatgttct 1140
ggactgttta ccttggaaga aggactcgca ggatttcttt ttgttttcaa actaaatgaa 1200
45 actatgttgt tgaatttgtt gggcaatgac ttctacatcc cccttttaag ctcaaataac 1260
tcaattgctg tgcaaaatga ggttcagtct gaagatgctc aggtacctga cagaagtaga 1320
gaaactaatt ttactgcata taccgatggt ataataagga aaataaggaa cttggtgagt 1380
gatatttctc ctgagaagag tcaaaggaaa agatccaaag aagcacaaga aaccattctt 1440
caagaaatag aaaaattggc tgctgaagcg tatagtatct tcagaactac tgttccaact 1500
50 ctaccggagg aaatcattgc agaaactgaa aaggtgaaag tcgcccctgc caaaatatgc 1560
tcggggacag gaacagggtt tgaaatatgt tgccaagggt ttaactggga atctagtaaa 1620
tctggaagat ggtacgagga actcaaaagt aaagctgcag aattatcctc actagggttc 1680
actgtgattt ggttcccacc tcctacagat tctgtgtcgc ctcaagggtg catgccgagg 1740
gatttataca atatgaactc cagatatgga aatatggacg aactgaagga gactgtgaag 1800
55 acattccatg atgccggttt aaaagtctct ggagatgctg ttctgaatca ccgttgtgca 1860
gaatatcaga atcaaaatgg tgtttggaat atatttggtg gtcgttttaa ttgggatgaa 1920
cgtgcagttg ttgcagatga tccacatttt cagggtaggg gcaacaaaag tagtggagat 1980
agttttcatg ctgccccaaa cattgatcat tcacaagatt ttgtgaggaa ggatatcaga 2040
gaatggttat gctggctaag ggacgatatt gggatgacg gatggaggct tgatttctgt 2100
60 agaggatttt ggggtggcta cgtcaaggac tacatggatg ccagtgagcc ctactttgcc 2160
gtaggcgagt attgggattc cctaagttat acatacgggg aaatggatca caatcaagat 2220
gcacacaggc agagaattgt tgattggatc aacgctacta atggaacttg tggatgcatt 2280
gatgtcacia caaaagggat tctccatgca gcactggaaa gatgcgagta ttggcgattg 2340

tcagatgaga agggtaagcc tccagggggtt cttgggatggg ggccatctcg tgctgtcact 2400
 ttcatagaga atcatgatac tggttctact caggggtcatt ggagggtttcc aaataaaaaa 2460
 gaaatgcaag gatatgcata catttttgact catccgggaa cacctacagt tttctatgac 2520
 cacattttct ctcactacca atctgaaatt gctgctctga tctctctcag aaaccggaac 2580
 5 aagctcaact gtcggagtag agttaaatac accaaggcag aaagagacgt ctacgcggcc 2640
 atcattgatg aaaagggtcgc catcaaaatc ggaccagggtc attacgaacc tgctagtggg 2700
 cctcaaaaatt ggaataaaaag ccttgaggga agagactaca aggtctggga agcgtcataa 2760
 attttttctg caccggtatg tatgtaactt cgagtgtata aggtttgagt accggatgat 2820
 10 tcaataagga agttcccata tttacattgt aagcaaatac aaggatgccg agatagagta 2880
 atatgtactt ctctctagag gtctggctga cacccttgaa ttgaagttgt ttcggcagac 2940
 attttccgaa ttttaaggac tgtagacgct caactgagga cgagacgggc cagcctgctt 3000
 cagcagcaca gaatcgcggc cagtcggctg agtgtggaag tgggtgagcat ggtactgggt 3060
 tccaagggca gtgctccatc ttacacacac tgttgactct ccaacttata ctgatgcttt 3120
 taaccaataa tgtataaatt ggtgcatgcc agtgactcgt ggtccatgaa cgttgagtga 3180
 15 gtttcctatc tgaactttta ttctgtctca cattgctcca gtccatcaac tttatgcatg 3240
 gactcgtaaa ttttttagta ttgttcttga aacaggatga tcgatatatc gttaatatgt 3300
 gatccatata ttatattaaa aaaa 3324

<210> 2
 20 <211> 901
 <212> PRT
 <213> Malus domestica

<400> 2
 25 Met Ser Thr Val Arg Ile Glu Pro Leu Leu His His Tyr Arg Arg Gln
 1 5 10 15
 Lys Pro Ser His Arg Leu Pro Pro Ser Lys His Pro Leu Lys Leu Ser
 20 25 30
 30 Ser Ser Phe Thr Ala Phe Pro Lys Lys Leu Val Val Ser Asn Gly Arg
 35 40 45
 Ser Phe Cys Asn Phe Glu Pro Pro Thr Leu Ser Val Arg Ala Ala Ser
 50 55 60
 Thr Asp Thr Ala Thr Val Glu Ala Thr Glu Phe Ala Asp Ala Phe Tyr
 65 70 75 80
 40 Lys Glu Thr Phe Pro Leu Lys Arg Thr Glu Val Val Glu Gly Lys Met
 85 90 95
 Ile Val Lys Leu Asp Asn Gly Lys Asp Ala Lys Asn Trp Val Leu Thr
 100 105 110
 45 Val Gly Cys Asn Leu Pro Gly Lys Trp Val Leu His Trp Gly Val Asn
 115 120 125
 Tyr Val Asp Asp Val Gly Ser Glu Trp Asp Gln Pro Pro Ser Glu Met
 130 135 140
 Arg Pro Ala Gly Ser Val Ser Ile Lys Asp Tyr Ala Ile Glu Thr Pro
 145 150 155 160
 55 Leu Lys Glu Ser Leu Ser Pro Val Gly Gly Asp Thr Ser His Glu Val
 165 170 175
 Lys Ile Asp Val Thr Pro Asn Ser Ala Ile Ala Ala Ile Asn Phe Val
 180 185 190
 60 Leu Lys Asp Glu Glu Thr Gly Ala Trp Tyr Gln His Arg Gly Arg Asp
 195 200 205

	Phe	Lys	Val	Pro	Phe	Val	Gly	Tyr	Leu	Gln	Asp	Asp	Asp	Asn	Val	Val	
	210						215					220					
5	Gly	Ala	Thr	Arg	Ala	Leu	Gly	Ala	Trp	Ser	Gly	Thr	Leu	Gly	Lys	Leu	
	225					230					235					240	
	Ser	Asn	Val	Phe	Val	Lys	Ala	Glu	Thr	Ser	Asn	Ser	Lys	Asp	Gln	Glu	
					245					250					255		
10	Ser	Ser	Ser	Glu	Ser	Arg	Asp	Pro	Gln	Gln	Lys	Thr	Met	Arg	Leu	Glu	
				260					265					270			
	Gly	Phe	Tyr	Glu	Glu	Leu	Pro	Ile	Ala	Lys	Glu	Ile	Ala	Val	Asn	His	
15			275					280					285				
	Ser	Ala	Thr	Val	Ser	Val	Arg	Lys	Cys	Pro	Glu	Thr	Thr	Lys	Asn	Leu	
	290						295					300					
20	Leu	Tyr	Leu	Glu	Thr	Asp	Leu	Pro	Asp	His	Ala	Val	Val	His	Trp	Gly	
	305					310					315					320	
	Val	Cys	Arg	Asp	Asp	Ala	Lys	Arg	Trp	Glu	Ile	Pro	Ala	Ala	Pro	His	
					325					330					335		
25	Pro	Pro	Glu	Thr	Val	Val	Phe	Lys	Asp	Lys	Ala	Leu	Arg	Thr	Arg	Leu	
				340					345					350			
	Gln	Gln	Arg	Glu	Asp	Gly	Asn	Gly	Cys	Ser	Gly	Leu	Phe	Thr	Leu	Glu	
30			355					360					365				
	Glu	Gly	Leu	Ala	Gly	Phe	Leu	Phe	Val	Phe	Lys	Leu	Asn	Glu	Thr	Met	
	370						375					380					
35	Trp	Leu	Asn	Cys	Val	Gly	Asn	Asp	Phe	Tyr	Ile	Pro	Leu	Leu	Ser	Ser	
	385					390					395					400	
	Asn	Asn	Ser	Ile	Ala	Val	Gln	Asn	Glu	Val	Gln	Ser	Glu	Asp	Ala	Gln	
				405						410					415		
40	Val	Pro	Asp	Arg	Ser	Arg	Glu	Thr	Asn	Phe	Thr	Ala	Tyr	Thr	Asp	Gly	
				420					425					430			
	Ile	Ile	Asn	Glu	Ile	Arg	Asn	Leu	Val	Ser	Asp	Ile	Ser	Ser	Glu	Lys	
45			435					440					445				
	Ser	Gln	Arg	Lys	Arg	Ser	Lys	Glu	Ala	Gln	Glu	Thr	Ile	Leu	Gln	Glu	
	450						455					460					
50	Ile	Glu	Lys	Leu	Ala	Ala	Glu	Ala	Tyr	Ser	Ile	Phe	Arg	Thr	Thr	Val	
	465					470					475					480	
	Pro	Thr	Leu	Pro	Glu	Glu	Ile	Ile	Ala	Glu	Thr	Glu	Lys	Val	Lys	Val	
					485					490					495		
55	Ala	Pro	Ala	Lys	Ile	Cys	Ser	Gly	Thr	Gly	Thr	Gly	Phe	Glu	Ile	Leu	
				500					505					510			
	Cys	Gln	Gly	Phe	Asn	Trp	Glu	Ser	Ser	Lys	Ser	Gly	Arg	Trp	Tyr	Glu	
60			515					520				525					
	Glu	Leu	Lys	Ser	Lys	Ala	Ala	Glu	Leu	Ser	Ser	Leu	Gly	Phe	Thr	Val	
	530						535					540					

	Ile	Trp	Phe	Pro	Pro	Pro	Thr	Asp	Ser	Val	Ser	Pro	Gln	Gly	Tyr	Met	
	545					550					555					560	
5	Pro	Arg	Asp	Leu	Tyr	Asn	Met	Asn	Ser	Arg	Tyr	Gly	Asn	Met	Asp	Glu	
					565					570					575		
	Leu	Lys	Glu	Thr	Val	Lys	Thr	Phe	His	Asp	Ala	Gly	Leu	Lys	Val	Leu	
				580					585					590			
10	Gly	Asp	Ala	Val	Leu	Asn	His	Arg	Cys	Ala	Glu	Tyr	Gln	Asn	Gln	Asn	
			595					600					605				
	Gly	Val	Trp	Asn	Ile	Phe	Gly	Gly	Arg	Leu	Asn	Trp	Asp	Glu	Arg	Ala	
15		610					615					620					
	Val	Val	Ala	Asp	Asp	Pro	His	Phe	Gln	Gly	Arg	Gly	Asn	Lys	Ser	Ser	
	625					630					635					640	
20	Gly	Asp	Ser	Phe	His	Ala	Ala	Pro	Asn	Ile	Asp	His	Ser	Gln	Asp	Phe	
					645					650					655		
	Val	Arg	Lys	Asp	Ile	Arg	Glu	Trp	Leu	Cys	Trp	Leu	Arg	Asp	Asp	Ile	
				660					665					670			
25	Gly	Tyr	Asp	Gly	Trp	Arg	Leu	Asp	Phe	Val	Arg	Gly	Phe	Trp	Gly	Gly	
			675					680					685				
	Tyr	Val	Lys	Asp	Tyr	Met	Asp	Ala	Ser	Glu	Pro	Tyr	Phe	Ala	Val	Gly	
30		690					695					700					
	Glu	Tyr	Trp	Asp	Ser	Leu	Ser	Tyr	Thr	Tyr	Gly	Glu	Met	Asp	His	Asn	
	705					710					715					720	
35	Gln	Asp	Ala	His	Arg	Gln	Arg	Ile	Val	Asp	Trp	Ile	Asn	Ala	Thr	Asn	
				725						730					735		
	Gly	Thr	Cys	Gly	Ala	Phe	Asp	Val	Thr	Thr	Lys	Gly	Ile	Leu	His	Ala	
				740					745					750			
40	Ala	Leu	Glu	Arg	Cys	Glu	Tyr	Trp	Arg	Leu	Ser	Asp	Glu	Lys	Gly	Lys	
			755					760					765				
	Pro	Pro	Gly	Val	Leu	Gly	Trp	Trp	Pro	Ser	Arg	Ala	Val	Thr	Phe	Ile	
45		770					775					780					
	Glu	Asn	His	Asp	Thr	Gly	Ser	Thr	Gln	Gly	His	Trp	Arg	Phe	Pro	Asn	
	785					790					795					800	
50	Lys	Lys	Glu	Met	Gln	Gly	Tyr	Ala	Tyr	Ile	Leu	Thr	His	Pro	Gly	Thr	
					805					810					815		
	Pro	Thr	Val	Phe	Tyr	Asp	His	Ile	Phe	Ser	His	Tyr	Gln	Ser	Glu	Ile	
				820					825					830			
55	Ala	Ala	Leu	Ile	Ser	Leu	Arg	Asn	Arg	Asn	Lys	Leu	Asn	Cys	Arg	Ser	
			835					840					845				
	Arg	Val	Lys	Ile	Thr	Lys	Ala	Glu	Arg	Asp	Val	Tyr	Ala	Ala	Ile	Ile	
60		850					855					860					
	Asp	Glu	Lys	Val	Ala	Ile	Lys	Ile	Gly	Pro	Gly	His	Tyr	Glu	Pro	Ala	
	865					870					875					880	

WO 03/027141

PCT/NZ02/00200

Ser Gly Pro Gln Asn Trp Asn Lys Ser Leu Glu Gly Arg Asp Tyr Lys
885 890 895

5 Val Trp Glu Ala Ser
900

<210> 3
<211> 1468
10 <212> DNA
<213> Malus domestica

<400> 3
15 agcggccggtt gctcgttcca tgtcgacggt taggatagag cccctcctcc accactaccg 60
tcgacagaaa cccagccacc gccttccgcc gtogaagcac ccattaaagc tcagctcttc 120
tttactgct tttccaaaga agctagtagt ctccaatggc cgcagcttct gcaacttcca 180
gcctcccact ctcaagtgtca gagctgcctc caccgataca gccaccgtcg aggccaccga 240
attcgccgac gccttctaca aggagacctt ccctctcaag cgaactgaag tgggtggagg 300
aaagatgatc gtgaaattag ataatgggaa ggatgcaaag aattgggtgc tgactgtggg 360
20 ttgtaatctt cctggaaaat gggttcttca ctggggaggt aattatgtgg atgacgtcgg 420
cagtgaatgg gatcagcctc ctagtgaaat gagaccagct ggttcagttt ccatcaagga 480
ctatgcaata gagacaccct tgaaggaatc gttgtcgccg gtgggaggcg atacatctca 540
cgaagtgaag attgatgtta caccacaacag cgcaatcgca gcgataaatt ttgttctcaa 600
ggatgaagaa accggtgcct ggtatcagca tagagggaga gactttaaaag tgcctttcgt 660
25 gggctacctg caagacgatg acaatgtagt tggagcaaca agggccttgg gcgcgtggtc 720
aggaactttg ggaaaactat ccaatgtgtt tgtcaaagca gaaacatcaa attccaaaga 780
tcaagaaagc agcagtgaat ccagagaccc tcaacagaaa actatgcgtc tagaaggatt 840
ctatgaagaa ctgccaattg caaaagaaat tgctgttaac cattcagcaa ctgtttccgt 900
taggaagtgc cctgagacta ctaagaatct tctatacttg gaaacagatt tacctgatca 960
30 tgctgttggt cactggggag tttgcagaga tgatgctaaa agatgggaaa ttccagctgc 1020
cccgcaccca ccagaaacag tagttttcaa ggacaaggct ttgcggactc gattacagca 1080
aarggaggat ggaaatggat gttctggact gtttaccttg raagaaggac tcgcaggatt 1140
tctttttgtt ttcaaaactaa atgaamytat gtggttgaaat tgtgtgggca atgacttcta 1200
catccccctt ttaagctcaa ataactcaat tgctgtgcaa aatgaggttc agtctgaaga 1260
35 tgctcaggta cctgmccagaa gtagagaaac taattttact gcatataaccg atggtataat 1320
caatgaaata aggaacttgg tgagtgatat ttcctctgag aagagtcaaa ggaaaagatc 1380
caaagaagca caagaaacca ttcttcaaga aatagaaaaa ttggctgctg aagcgtatag 1440
tatcttcaga actactgttc caactcta 1468

40 <210> 4
<211> 483
<212> PRT
<213> Malus domestica

45 <400> 4
Met Ser Thr Val Arg Ile Glu Pro Leu Leu His His Tyr Arg Arg Gln
1 5 10 15

50 Lys Pro Ser His Arg Leu Pro Pro Ser Lys His Pro Leu Lys Leu Ser
20 25 30

Ser Ser Phe Thr Ala Phe Pro Lys Lys Leu Val Val Ser Asn Gly Arg
35 40 45

55 Ser Phe Cys Asn Phe Gln Pro Pro Thr Leu Ser Val Arg Ala Ala Ser
50 55 60

60 Thr Asp Thr Ala Thr Val Glu Ala Thr Glu Phe Ala Asp Ala Phe Tyr
65 70 75 80

Lys Glu Thr Phe Pro Leu Lys Arg Thr Glu Val Val Glu Gly Lys Met
85 90 95

	Ile	Val	Lys	Leu	Asp	Asn	Gly	Lys	Asp	Ala	Lys	Asn	Trp	Val	Leu	Thr	
				100					105					110			
5	Val	Gly	Cys	Asn	Leu	Pro	Gly	Lys	Trp	Val	Leu	His	Trp	Gly	Val	Asn	
			115					120					125				
	Tyr	Val	Asp	Asp	Val	Gly	Ser	Glu	Trp	Asp	Gln	Pro	Pro	Ser	Glu	Met	
		130					135					140					
10	Arg	Pro	Ala	Gly	Ser	Val	Ser	Ile	Lys	Asp	Tyr	Ala	Ile	Glu	Thr	Pro	
		145				150					155					160	
	Leu	Lys	Glu	Ser	Leu	Ser	Pro	Val	Gly	Gly	Asp	Thr	Ser	His	Glu	Val	
					165					170					175		
15	Lys	Ile	Asp	Val	Thr	Pro	Asn	Ser	Ala	Ile	Ala	Ala	Ile	Asn	Phe	Val	
				180					185					190			
	Leu	Lys	Asp	Glu	Glu	Thr	Gly	Ala	Trp	Tyr	Gln	His	Arg	Gly	Arg	Asp	
20			195					200					205				
	Phe	Lys	Val	Pro	Phe	Val	Gly	Tyr	Leu	Gln	Asp	Asp	Asp	Asn	Val	Val	
		210					215					220					
25	Gly	Ala	Thr	Arg	Ala	Leu	Gly	Ala	Trp	Ser	Gly	Thr	Leu	Gly	Lys	Leu	
		225				230					235					240	
	Ser	Asn	Val	Phe	Val	Lys	Ala	Glu	Thr	Ser	Asn	Ser	Lys	Asp	Gln	Glu	
					245					250					255		
30	Ser	Ser	Ser	Glu	Ser	Arg	Asp	Pro	Gln	Gln	Lys	Thr	Met	Arg	Leu	Glu	
				260					265					270			
	Gly	Phe	Tyr	Glu	Glu	Leu	Pro	Ile	Ala	Lys	Glu	Ile	Ala	Val	Asn	His	
35			275					280					285				
	Ser	Ala	Thr	Val	Ser	Val	Arg	Lys	Cys	Pro	Glu	Thr	Thr	Lys	Asn	Leu	
		290					295					300					
40	Leu	Tyr	Leu	Glu	Thr	Asp	Leu	Pro	Asp	His	Ala	Val	Val	His	Trp	Gly	
		305				310					315					320	
	Val	Cys	Arg	Asp	Asp	Ala	Lys	Arg	Trp	Glu	Ile	Pro	Ala	Ala	Pro	His	
					325					330					335		
45	Pro	Pro	Glu	Thr	Val	Val	Phe	Lys	Asp	Lys	Ala	Leu	Arg	Thr	Arg	Leu	
					340				345					350			
	Gln	Gln	Xaa	Glu	Asp	Gly	Asn	Gly	Cys	Ser	Gly	Leu	Phe	Thr	Leu	Xaa	
50			355					360					365				
	Glu	Gly	Leu	Ala	Gly	Phe	Leu	Phe	Val	Phe	Lys	Leu	Asn	Glu	Xaa	Met	
		370					375					380					
55	Trp	Leu	Asn	Cys	Val	Gly	Asn	Asp	Phe	Tyr	Ile	Pro	Leu	Leu	Ser	Ser	
		385				390					395					400	
	Asn	Asn	Ser	Ile	Ala	Val	Gln	Asn	Glu	Val	Gln	Ser	Glu	Asp	Ala	Gln	
					405					410					415		
60	Val	Pro	Xaa	Arg	Ser	Arg	Glu	Thr	Asn	Phe	Thr	Ala	Tyr	Thr	Asp	Gly	
				420					425						430		

Ile Ile Asn Glu Ile Arg Asn Leu Val Ser Asp Ile Ser Ser Glu Lys
 435 440 445
 5 Ser Gln Arg Lys Arg Ser Lys Glu Ala Gln Glu Thr Ile Leu Gln Glu
 450 455 460
 Ile Glu Lys Leu Ala Ala Glu Ala Tyr Ser Ile Phe Arg Thr Thr Val
 465 470 475 480
 10 Pro Thr Leu
 15 <210> 5
 <211> 265
 <212> PRT
 <213> Malus domestica
 20 <400> 5
 Trp Val Leu His Trp Gly Val Asn Tyr Val Asp Asp Val Gly Ser Glu
 1 5 10 15
 Trp Asp Gln Pro Pro Ser Glu Met Arg Pro Ala Gly Ser Val Ser Ile
 20 25 30
 25 Lys Asp Tyr Ala Ile Glu Thr Pro Leu Lys Glu Ser Leu Ser Pro Val
 35 40 45
 30 Gly Gly Asp Thr Ser His Glu Val Lys Ile Asp Val Thr Pro Asn Ser
 50 55 60
 Ala Ile Ala Ala Ile Asn Phe Val Leu Lys Asp Glu Glu Thr Gly Ala
 65 70 75 80
 35 Trp Tyr Gln His Arg Gly Arg Asp Phe Lys Val Pro Phe Val Gly Tyr
 85 90 95
 Leu Gln Asp Asp Asp Asn Val Val Gly Ala Thr Arg Ala Leu Gly Ala
 100 105 110
 40 Trp Ser Gly Thr Leu Gly Lys Leu Ser Asn Val Phe Val Lys Ala Glu
 115 120 125
 Thr Ser Asn Ser Lys Asp Gln Glu Ser Ser Ser Glu Ser Arg Asp Pro
 130 135 140
 Gln Gln Lys Thr Met Arg Leu Glu Gly Phe Tyr Glu Glu Leu Pro Ile
 145 150 155 160
 50 Ala Lys Glu Ile Ala Val Asn His Ser Ala Thr Val Ser Val Arg Lys
 165 170 175
 Cys Pro Glu Thr Thr Lys Asn Leu Leu Tyr Leu Glu Thr Asp Leu Pro
 180 185 190
 55 Asp His Ala Val Val His Trp Gly Val Cys Arg Asp Asp Ala Lys Arg
 195 200 205
 Trp Glu Ile Pro Ala Ala Pro His Pro Pro Glu Thr Val Val Phe Lys
 210 215 220
 60 Asp Lys Ala Leu Arg Thr Arg Leu Gln Gln Xaa Glu Asp Gly Asn Gly
 225 230 235 240

WO 03/027141

PCT/NZ02/00200

Cys Ser Gly Leu Phe Thr Leu Xaa Glu Gly Leu Ala Gly Phe Leu Phe
245 250 255

5 Val Phe Lys Leu Asn Glu Xaa Met Trp
260 265

10 <210> 6
<211> 183
<212> DNA
<213> Malus domestica

15 <400> 6
atgtcgacgg ttaggataga gccctcctc caccactacc gtgcacagaa acccagccac 60
cgccttccgc cgtcgaagca cccattaaag ctccagctctt ctttcactgc ttttccaaag 120
aagctagtag tctccaatgg ccgcagcttc tgcaacttcc agcctccac tctcagtgtc 180
aga 183

20 <210> 7
<211> 61
<212> PRT
<213> Malus domestica

25 <400> 7
Met Ser Thr Val Arg Ile Glu Pro Leu Leu His His Tyr Arg Arg Gln
1 5 10 15

30 Lys Pro Ser His Arg Leu Pro Pro Ser Lys His Pro Leu Lys Leu Ser
20 25 30

Ser Ser Phe Thr Ala Phe Pro Lys Lys Leu Val Val Ser Asn Gly Arg
35 40 45

35 Ser Phe Cys Asn Phe Gln Pro Pro Thr Leu Ser Val Arg
50 55 60

40 <210> 8
<211> 55
<212> PRT
<213> Arabidopsis thaliana

45 <400> 8
Met Ser Thr Val Pro Ile Glu Ser Leu Leu His His Ser Tyr Leu Arg
1 5 10 15

His Asn Ser Lys Val Asn Arg Gly Asn Arg Ser Phe Ile Pro Ile Ser
20 25 30

50 Leu Asn Leu Arg Ser His Phe Thr Ser Asn Lys Leu Leu His Ser Ile
35 40 45

55 Gly Lys Ser Val Gly Val Ser
50 55

60 <210> 9
<211> 934
<212> DNA
<213> Malus domestica

<400> 9

```

gtttctcagtc tmccgagtgg ccggttgctca taccatgtcg acggtttagga tagagcccct 60
cctccaggac ctccaccact acggtcgaca aaaaccagc caccgccgtc cgcagtcgaa 120
tcatccatta aagctcagct cttctttcac tgcttttcct aagaagctag tagtctccaa 180
cagccgcagc ttctgttact tccagcctcc cactcccagg aggggtccca ctctcggagt 240
5 cagagctgcc tccaccgata caaccactgt cgagacctcc gaatccaccg atcccatcta 300
caagaagacc ttccctctca agcggacgga agtgggtggag ggaaagatct ttgtgaaatt 360
agatcatggg aagaatgaaa agaagtgggt gctgactgtg ggttgtaatc ttcctggaaa 420
atgggttctt cattggggag tttcttttgt ggacgatgtt agttgcgaat gggaacagcc 480
tcctagttaa atgagaccag ctgggttcagt tcccatcaag gactatgcaa tagagacacc 540
10 cttgaaggaa tcgttgctcg ctgtgggagg cgatacatct tatgaagtga agattgatgt 600
taaaccacac agcgcaattg cagcgataaa ttttggttctc aaggatgaag aaactgggtgc 660
ttgggtatcag cacagaggga gtgacttttag agtgcctctt gtcgcctacc cccaagacga 720
tgacaatgta gttggagcga caaagggtct tggcatgtgg ccagggttgc aatgccactc 780
tattctctgg tttgcttgat ttattttgtt tgtaatactg ktaaatttga accaatttgt 840
15 gattttccaa gaaacggtgt aataaaagcc caaaaggrgm ctttcctaata tcagcttaaa 900
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaa 934

```

```

<210> 10
<211> 254
20 <212> PRT
    <213> Malus domestica

```

```

<400> 10
25 Met Ser Thr Val Arg Ile Glu Pro Leu Leu Gln Asp Leu His His Tyr
    1          5          10          15

Gly Arg Gln Lys Pro Ser His Arg Arg Pro Gln Ser Asn His Pro Leu
          20          25          30

30 Lys Leu Ser Ser Ser Phe Thr Ala Phe Pro Lys Lys Leu Val Val Ser
    35          40          45

Asn Ser Arg Ser Phe Cys Tyr Phe Gln Pro Pro Thr Pro Arg Arg Gly
    50          55          60

35 Pro Thr Leu Gly Val Arg Ala Ala Ser Thr Asp Thr Thr Thr Val Glu
    65          70          75          80

Thr Ser Glu Ser Thr Asp Pro Ile Tyr Lys Lys Thr Phe Pro Leu Lys
40          85          90          95

Arg Thr Glu Val Val Glu Gly Lys Ile Phe Val Lys Leu Asp His Gly
          100          105          110

45 Lys Asn Glu Lys Lys Trp Val Leu Thr Val Gly Cys Asn Leu Pro Gly
    115          120          125

Lys Trp Val Leu His Trp Gly Val Ser Phe Val Asp Asp Val Ser Cys
50          130          135          140

Glu Trp Glu Gln Pro Pro Ser Glu Met Arg Pro Ala Gly Ser Val Pro
145          150          155          160

Ile Lys Asp Tyr Ala Ile Glu Thr Pro Leu Lys Glu Ser Leu Ser Ser
55          165          170          175

Val Gly Gly Asp Thr Ser Tyr Glu Val Lys Ile Asp Val Lys Pro Asn
          180          185          190

60 Ser Ala Ile Ala Ala Ile Asn Phe Val Leu Lys Asp Glu Glu Thr Gly
    195          200          205

Ala Trp Tyr Gln His Arg Gly Ser Asp Phe Arg Val Pro Leu Val Ala

```

210 215 220

Tyr Pro Gln Asp Asp Asp Asn Val Val Gly Ala Thr Lys Gly Phe Gly
 225 230 235 240

5 Met Trp Pro Gly Cys Gln Cys His Ser Ile Leu Trp Phe Ala
 245 250

10 <210> 11
 <211> 519
 <212> DNA
 <213> Malus domestica

15 <400> 11
 gaaactaaat gaaagtacgt gggtgagatg tgtgggaaat gacttctaca tccccctttc 60
 aagctcaaaa aacgcaaagtg ttgtacaatc agagattcaa tctaaagatg ctcaggtacc 120
 tgacgggagt acagaagcag tagaagaaag tactgcatat gccgatggat taatcaatga 180
 aatgaggaac ttggtgagtg acgttttctc tgataaaaagt ccaaggacaa gatccaaaaa 240
 20 agcacaagaa gccattcttc aagaaataga aaaattggct gctgaagcgt atagtatctt 300
 cagaactacc gttccaactt tacctgagga aaccattgca gaaactgaag aagtgaaggt 360
 cgccccctgcc aaaatatgtt cagggacagg aacagggttc gaaatattgt gccagggttt 420
 taactgggaa tctagtaaag ctggaagatg gtacatggaa ctcaaaagta aagctgcact 480
 gttatcttca ctaggtttca ctgtgatttg gttcccacc 519

25 <210> 12
 <211> 172
 <212> PRT
 <213> Malus domestica

30 <400> 12
 Lys Leu Asn Glu Ser Thr Trp Leu Arg Cys Val Gly Asn Asp Phe Tyr
 1 5 10 15

35 Ile Pro Leu Ser Ser Ser Lys Asn Ala Asn Val Val Gln Ser Glu Ile
 20 25 30

Gln Ser Lys Asp Ala Gln Val Pro Asp Gly Ser Thr Glu Ala Val Glu
 35 40 45

40 Glu Ser Thr Ala Tyr Ala Asp Gly Leu Ile Asn Glu Met Arg Asn Leu
 50 55 60

45 Val Ser Asp Val Phe Ser Asp Lys Ser Pro Arg Thr Arg Ser Lys Lys
 65 70 75 80

Ala Gln Glu Ala Ile Leu Gln Glu Ile Glu Lys Leu Ala Ala Glu Ala
 85 90 95

50 Tyr Ser Ile Phe Arg Thr Thr Val Pro Thr Leu Pro Glu Glu Thr Ile
 100 105 110

Ala Glu Thr Glu Glu Val Lys Val Ala Pro Ala Lys Ile Cys Ser Gly
 115 120 125

55 Thr Gly Thr Gly Phe Glu Ile Leu Cys Gln Gly Phe Asn Trp Glu Ser
 130 135 140

60 Ser Lys Ser Gly Arg Trp Tyr Met Glu Leu Lys Ser Lys Ala Ala Leu
 145 150 155 160

Leu Ser Ser Leu Gly Phe Thr Val Ile Trp Phe Pro
 165 170

WO 03/027141

PCT/NZ02/00200

<210> 13
 <211> 426
 5 <212> DNA
 <213> *Malus domestica*

<400> 13
 10 gatatcagag aatgggttatg ctggctaagg aacgatattg ggtacgatgg atggaggctg 60
 gatttttgta gaggatttcg ggggtggctat gtcagggact acgtggatgc tagtgagccc 120
 tactttgctg taggcgagta ttgggattcc ctaagttata cgtatgggga aatggatcgc 180
 aatcaagatg cacacaggca gagaattgtt gattggatca acgctactaa tggaacttgt 240
 ggtgcatttg atgtcacaac aaaagggatt ctccatgcag cgctggaaag gtgcgagtat 300
 15 tggcgattgt cagatgagaa gggtaagcct ccaggggttc ttggatgggtg gccatctcgt 360
 gctgccactt tcatagagaa tcatgatact ggttctactc agggtcattg gaggtttcca 420
 aataaa 426

<210> 14
 <211> 142
 20 <212> PRT
 <213> *Malus domestica*

<400> 14
 25 Asp Ile Arg Glu Trp Leu Cys Trp Leu Arg Asn Asp Ile Gly Tyr Asp
 1 5 10 15
 Gly Trp Arg Leu Asp Phe Val Arg Gly Phe Arg Gly Gly Tyr Val Arg
 20 25 30
 30 Asp Tyr Val Asp Ala Ser Glu Pro Tyr Phe Ala Val Gly Glu Tyr Trp
 35 40 45
 Asp Ser Leu Ser Tyr Thr Tyr Gly Glu Met Asp Arg Asn Gln Asp Ala
 50 55 60
 35 His Arg Gln Arg Ile Val Asp Trp Ile Asn Ala Thr Asn Gly Thr Cys
 65 70 75 80
 40 Gly Ala Phe Asp Val Thr Thr Lys Gly Ile Leu His Ala Ala Leu Glu
 85 90 95
 Arg Cys Glu Tyr Trp Arg Leu Ser Asp Glu Lys Gly Lys Pro Pro Gly
 100 105 110
 45 Val Leu Gly Trp Trp Pro Ser Arg Ala Ala Thr Phe Ile Glu Asn His
 115 120 125
 Asp Thr Gly Ser Thr Gln Gly His Trp Arg Phe Pro Asn Lys
 130 135 140
 50

<210> 15
 <211> 2664
 <212> DNA
 55 <213> *Arabidopsis thaliana*

<400> 15
 60 atgtccactg ttcccattga gtctcttctc caccattctt atcttcgtca caactcaaaa 60
 gtcaatcgcg gaaatagaag cttcatacct atctcgttga atctccgttc tcatttcact 120
 tctaacaaac tactacactc aattggaaaa agtgctcggtg ttagctcgat gaacaaaagt 180
 cccgtcgcca ttcgtgctac ctcatcggat actgccgtcg tggaaaccgc tcaatcggat 240
 gatgtcatct tcaaggaaat tttccctgtt cagcgaatcg aaaaggcaga aggaaagatt 300
 tatgttcgat taaaggaagt gaaggagaag aattgggagc tgagtgttgg atgtagtata 360

	cctggaaaat	ggatTTTTgca	ttgggggagtt	tcatatgtgg	gtgatactgg	cagtgaatgg	420
	gatcaacctc	cgaagatat	gagacctcct	ggctcaattg	ccattaagga	ctatgccata	480
	gagacacctt	tgaagaagtt	atctgaagga	gattcttttt	ttgaagttgc	tattaatcta	540
5	aatctggaga	gttctgtagc	agctctcaat	tttgttttga	aggacgaaga	aactggggcg	600
	tgggtatcagc	acaaagggag	agactttaag	gttcctcttg	tagacgatgt	tcctgataat	660
	ggaaatttga	tcggagctaa	gaaaggattt	ggtgcccttg	ggcagctttc	aaacatccct	720
	ctcaaacaag	ataagtccag	tgcagagact	gattctattg	aagaaagaaa	aggtcttcaa	780
	gagtttttacg	aggagatgcc	aataagtaaa	cgtgttgccg	atgataactc	agtcagtgtc	840
10	actgcaagga	aatgccctga	aacatctaag	aacattgtat	caattgaaac	cgattttaccg	900
	ggtgacgtta	ctgtccactg	gggagtttgc	aaaaacggca	ctaagaaatg	ggaaattcca	960
	tctgaacctt	accctgaaga	gacatctttg	tttaagaaca	aggcattacg	cactcgttta	1020
	cagcgaaaag	acgatggaaa	tggatcgttt	ggattattct	ctctggatgg	aaagcttgaa	1080
	gggttatgct	ttgttcttaa	gttaaataag	aatacttggc	taaattatag	gggggaagac	1140
15	ttctatgtcc	ctttccttac	ttcaagtagc	tcgcccgttg	aaactgaagc	tgcccaagtg	1200
	agtaaacccta	aacgaaaaac	agataaagaa	gtgtctgcta	gtggatttac	taaagaaatc	1260
	atcacggaga	taaggaactt	ggcaattgac	atttcctctc	ataagaatca	gaagacaaac	1320
	gtcaaagaag	tgcaggaaaa	cattctacaa	gaaattgaga	aactggctgc	ggaggcatac	1380
	agcatattta	gaagcacaa	tccagctttt	tccgaggaag	gtgttttaga	agcagaagct	1440
	gacaagcctg	acattaaaat	ctcctcagga	accggctcgg	gatttgagat	attatgccaa	1500
20	ggtttcaact	gggagtccta	taaactctgg	agatgggtact	tggaaactta	agaaaaagcc	1560
	gatgagttag	cttcacttgg	attcactgtt	ctgtgggtac	ctccaccgac	agaatctgtg	1620
	tcacctgaag	ggtacatgcc	taaagacctg	tataacttga	attccagata	tggacaatt	1680
	gatgagctaa	aagatacagt	gaagaaattt	cacaaggtcg	ggatcaaagt	tttaggcgat	1740
	gctgttctga	atcacccgtg	tgcacacttc	aaaaatcaaa	atggtgtatg	gaatctattt	1800
25	ggaggacgtc	taaactggga	cgatagggca	gtagttgcag	atgacctca	tttccagggt	1860
	agaggaaaca	agagcagtgg	agataatttc	catgctgctc	caaacataga	tcactcgcaa	1920
	gactttgtta	ggaaggatat	caaggaatgg	ctatgctgga	tgatggaaga	agttgggtat	1980
	gatggatgga	ggcttgactt	tgtaaagagg	ttctggggag	gttatgtgaa	agactatatg	2040
	gatgctagca	aaccgtactt	tgcgggttgg	gaatattggg	attcgtttaag	ttacacgtac	2100
30	ggagaaatgg	actacaatca	agacgcacat	cgtcaaagaa	tagttgactg	gattaatgca	2160
	actagtggag	ctgctgggtc	atttgatgtc	actaccaaag	gaattcttca	tacggcgctt	2220
	caaaaatgtg	aatattggag	actctcagac	ccaaaaggga	agcctccagg	tgtagttggg	2280
	tgggtggcctt	ctcgtgctgt	aacatttcac	gagaatcatg	acactggctc	tacacagggt	2340
	cattggagat	ttccggaggg	gaaggaaatg	caaggatatg	cttacatcct	aactcatcca	2400
35	ggaacaccag	cggctcttct	cgaccatata	ttctcggtat	atcattccga	gattgctgca	2460
	cttctctctc	tcaggaacag	acagaaactc	cactgtcgga	gtgaggtgaa	tatagacaag	2520
	agtgagagag	atgtgtatgc	ggctataata	gatgaaaagg	ttgcaatgaa	gatcggacca	2580
	gggcattatg	aaccacaaaa	cggatcgcaa	aactggtctg	tagccgttga	aggcagagac	2640
40	tacaaggtgt	gggaaacatc	ttaa				2664
	<210> 16						
	<211> 2939						
	<212> DNA						
	<213> Arabidopsis thaliana						
45	<400> 16						
	aaaatatctt	cttcttcttt	cttctgctct	ctctctctct	ctcgccatgt	ccactgttcc	60
	cattgagtct	cttctccacc	attcttatct	tcgtcacaac	tcaaaagtca	atcgcgga	120
	tagaagcttc	atacctatct	cgttgaatct	ccgttctcat	ttcacttcta	acaaactact	180
50	acactcaatt	ggaaaaagtg	tcgggtgtag	ctcgatgaac	aaaagtcccg	tcgccattcg	240
	tgtacacctc	tcggatactg	ccgtcgtgga	aaccgctcaa	tcggatgatg	tcatcttcaa	300
	ggaaattttc	cctgttcagc	gaatcgaaaa	ggcagaagga	aagatttatg	ttcgattaaa	360
	ggaagtgaag	gagaagaatt	gggagctgag	tgttggtatg	agtataacctg	gaaaatggat	420
	tttgcattgg	ggagtttcat	atgtgggtga	tactggcagt	gaatgggatc	aacctccgga	480
55	agatatgaga	cctcctggct	caattgccat	taaggactat	gccatagaga	cacctttgaa	540
	gaagttatct	gaaggagatt	ctttttttga	agttgctatt	aatctaaatc	tggagagttc	600
	tgtagcagct	ctcaattttg	ttttgaagga	cgaagaaact	ggggcggtgt	atcagcacia	660
	aggagagagc	tttaagggtc	ctcttgtaga	cgatgttcct	gataatggaa	atttgatcgg	720
	agctaagaaa	ggatttggtg	cccttgggca	gctttcaaac	atccctctca	aacaagataa	780
60	gtccagtgca	gagactgatt	ctattgaaga	aagaaaagg	cttcaagagt	tttacgagga	840
	gatgccataa	agtaaaccgtg	ttgccgatga	taactcagtc	agtgtcactg	caaggaaatg	900
	ccctgaaaca	tctaagaaca	ttgtatcaat	tgaaaccgat	ttaccgggtg	acgttactgt	960
	ccactgggga	gtttgcaaaa	acggcactaa	gaaatgggaa	attccatctg	aaccttacct	1020

	tgaagagaca	tctttgttta	agaacaaggc	attacgcact	cgtttacagc	gaaaagacga	1080									
	tggaaatgga	tcgtttggat	tattctctct	ggatggaaag	cttgaagggg	tatgctttgt	1140									
	tcttaagtta	aatgaaaata	cttggctaaa	ttataggggg	gaagacttct	atgtcccttt	1200									
	ccttacttca	agtagctcgc	ccgttgaaac	tgaagctgcc	caagtgaagta	aacccaaaacg	1260									
5	aaaaacagat	aaagaagtgt	ctgctagtgg	atttactaaa	gaaatcatca	cggagataag	1320									
	gaacttggca	attgacattt	cctctcataa	gaatcagaag	acaaacgtca	aagaagtgca	1380									
	ggaaaacatt	ctacaagaaa	ttgagaaact	ggctgcggag	gcatatagca	tatttagaag	1440									
	cacaactcca	gctttttccg	aggaaggtgt	tttagaagca	gaagctgaca	agcctgacat	1500									
	taaaatctcc	tcaggaaccg	gctcgggatt	tgagatatta	tgccaagggt	tcaactggga	1560									
10	gtccaataaa	tctgggagat	ggtacttggg	acttcaagaa	aaagccgatg	agttagcttc	1620									
	acttggattc	actgttctgt	ggttacctcc	accgacagaa	tctgtgtcac	ctgaagggta	1680									
	catgcctaaa	gacctgtata	acttgaattc	cagatatgga	acaattgatg	agctaaaaga	1740									
	tacagtgaag	aaatttcaca	aggtcgggat	caaagtttta	ggcgatgctg	ttctgaatca	1800									
	ccgctgtgca	cacttcaaaa	atcaaaatgg	tgtatggaat	ctatttggag	gacgtctaaa	1860									
15	ctgggacgat	agggcagtag	ttgcagatga	ccctcatttc	cagggtagag	gaaacaagag	1920									
	cagtggagat	aatttccatg	ctgctccaaa	catagatcac	tcgcaagact	ttgttaggaa	1980									
	ggatatcaag	gaatggctat	gctggatgat	ggaagaagtt	gggtatgatg	gatggaggct	2040									
	tgactttgta	agagggttct	ggggagggtta	tgtgaaagac	tatatggatg	ctagcaaacc	2100									
	gtactttgcg	gttggtgaat	attgggattc	gttaagttac	acgtacggag	aatggacta	2160									
20	caatcaagac	gcacatcgtc	aaagaatagt	tgactggatt	aatgcaacta	gtggagctgc	2220									
	tggtgcattt	gatgtcacta	ccaaaggaat	tcttcatacg	gcgcttcaaa	aatgtgaata	2280									
	ttggagactc	tcagacccaa	aagggaagcc	tccaggtgta	gttggttggt	ggccttctcg	2340									
	tgctgtaaca	ttcatcgaga	atcatgacac	tggctctaca	cagggtcatt	ggagatttcc	2400									
	ggaggggaag	gaaatgcaag	gatatgctta	catcctaact	catccaggaa	caccagcggg	2460									
25	cttcttcgac	catatcttct	cggattatca	ttccgagatt	gctgcacttc	tctctctcag	2520									
	gaacagacag	aaactccact	gtcggagtga	ggtgaatata	gacaagagtg	agagagatgt	2580									
	gtatgcggct	ataatagatg	aaaaggttgc	aatgaagatc	ggaccagggc	attatgaacc	2640									
	accaaacgga	tcgcaaaaact	ggtctgtagc	cgttgaaggc	agagactaca	aggtgtggga	2700									
	aacatcttaa	gaagaagggt	cccaattttt	atgtgtatct	agtagtaa	aaaattgttc	2760									
30	ctggaggata	tattatttca	gggaataaga	ttacctggaa	atgttatatt	tagagtggaa	2820									
	ttaaaaccag	atctaagatt	tgtcccactt	ttttcaatgc	tgagagaata	ttcatattgt	2880									
	tgtattggat	gcttgtaaag	tgtatgcata	tatatatgaa	taataatata	tctatttcc	2939									
	<210> 17															
35	<211> 887															
	<212> PRT															
	<213> Arabidopsis thaliana															
	<400> 17															
40	Met	Ser	Thr	Val	Pro	Ile	Glu	Ser	Leu	Leu	His	His	Ser	Tyr	Leu	Arg
	1				5				10						15	
	His	Asn	Ser	Lys	Val	Asn	Arg	Gly	Asn	Arg	Ser	Phe	Ile	Pro	Ile	Ser
				20					25					30		
45	Leu	Asn	Leu	Arg	Ser	His	Phe	Thr	Ser	Asn	Lys	Leu	Leu	His	Ser	Ile
			35					40				45				
	Gly	Lys	Ser	Val	Gly	Val	Ser	Ser	Met	Asn	Lys	Ser	Pro	Val	Ala	Ile
50		50					55					60				
	Arg	Ala	Thr	Ser	Ser	Asp	Thr	Ala	Val	Val	Glu	Thr	Ala	Gln	Ser	Asp
	65					70					75				80	
55	Asp	Val	Ile	Phe	Lys	Glu	Ile	Phe	Pro	Val	Gln	Arg	Ile	Glu	Lys	Ala
					85				90					95		
	Glu	Gly	Lys	Ile	Tyr	Val	Arg	Leu	Lys	Glu	Val	Lys	Glu	Lys	Asn	Trp
				100					105					110		
60	Glu	Leu	Ser	Val	Gly	Cys	Ser	Ile	Pro	Gly	Lys	Trp	Ile	Leu	His	Trp
				115				120					125			

	Gly	Val	Ser	Tyr	Val	Gly	Asp	Thr	Gly	Ser	Glu	Trp	Asp	Gln	Pro	Pro
	130						135					140				
5	Glu	Asp	Met	Arg	Pro	Pro	Gly	Ser	Ile	Ala	Ile	Lys	Asp	Tyr	Ala	Ile
	145					150					155					160
	Glu	Thr	Pro	Leu	Lys	Lys	Leu	Ser	Glu	Gly	Asp	Ser	Phe	Phe	Glu	Val
					165					170					175	
10	Ala	Ile	Asn	Leu	Asn	Leu	Glu	Ser	Ser	Val	Ala	Ala	Leu	Asn	Phe	Val
				180					185					190		
	Leu	Lys	Asp	Glu	Glu	Thr	Gly	Ala	Trp	Tyr	Gln	His	Lys	Gly	Arg	Asp
15			195					200					205			
	Phe	Lys	Val	Pro	Leu	Val	Asp	Asp	Val	Pro	Asp	Asn	Gly	Asn	Leu	Ile
	210						215					220				
20	Gly	Ala	Lys	Lys	Gly	Phe	Gly	Ala	Leu	Gly	Gln	Leu	Ser	Asn	Ile	Pro
	225					230					235					240
	Leu	Lys	Gln	Asp	Lys	Ser	Ser	Ala	Glu	Thr	Asp	Ser	Ile	Glu	Glu	Arg
					245					250					255	
25	Lys	Gly	Leu	Gln	Glu	Phe	Tyr	Glu	Glu	Met	Pro	Ile	Ser	Lys	Arg	Val
				260					265					270		
	Ala	Asp	Asp	Asn	Ser	Val	Ser	Val	Thr	Ala	Arg	Lys	Cys	Pro	Glu	Thr
30			275					280					285			
	Ser	Lys	Asn	Ile	Val	Ser	Ile	Glu	Thr	Asp	Leu	Pro	Gly	Asp	Val	Thr
	290						295					300				
35	Val	His	Trp	Gly	Val	Cys	Lys	Asn	Gly	Thr	Lys	Lys	Trp	Glu	Ile	Pro
	305					310					315					320
	Ser	Glu	Pro	Tyr	Pro	Glu	Glu	Thr	Ser	Leu	Phe	Lys	Asn	Lys	Ala	Leu
					325					330					335	
40	Arg	Thr	Arg	Leu	Gln	Arg	Lys	Asp	Asp	Gly	Asn	Gly	Ser	Phe	Gly	Leu
				340					345					350		
	Phe	Ser	Leu	Asp	Gly	Lys	Leu	Glu	Gly	Leu	Cys	Phe	Val	Leu	Lys	Leu
45			355					360					365			
	Asn	Glu	Asn	Thr	Trp	Leu	Asn	Tyr	Arg	Gly	Glu	Asp	Phe	Tyr	Val	Pro
	370						375					380				
50	Phe	Leu	Thr	Ser	Ser	Ser	Ser	Pro	Val	Glu	Thr	Glu	Ala	Ala	Gln	Val
	385					390					395					400
	Ser	Lys	Pro	Lys	Arg	Lys	Thr	Asp	Lys	Glu	Val	Ser	Ala	Ser	Gly	Phe
					405					410					415	
55	Thr	Lys	Glu	Ile	Ile	Thr	Glu	Ile	Arg	Asn	Leu	Ala	Ile	Asp	Ile	Ser
				420					425					430		
	Ser	His	Lys	Asn	Gln	Lys	Thr	Asn	Val	Lys	Glu	Val	Gln	Glu	Asn	Ile
60			435					440					445			
	Leu	Gln	Glu	Ile	Glu	Lys	Leu	Ala	Ala	Glu	Ala	Tyr	Ser	Ile	Phe	Arg
	450						455					460				

	Ser	Thr	Thr	Pro	Ala	Phe	Ser	Glu	Glu	Gly	Val	Leu	Glu	Ala	Glu	Ala	
	465					470					475					480	
5	Asp	Lys	Pro	Asp	Ile	Lys	Ile	Ser	Ser	Gly	Thr	Gly	Ser	Gly	Phe	Glu	
					485					490					495		
	Ile	Leu	Cys	Gln	Gly	Phe	Asn	Trp	Glu	Ser	Asn	Lys	Ser	Gly	Arg	Trp	
				500					505					510			
10	Tyr	Leu	Glu	Leu	Gln	Glu	Lys	Ala	Asp	Glu	Leu	Ala	Ser	Leu	Gly	Phe	
			515					520					525				
	Thr	Val	Leu	Trp	Leu	Pro	Pro	Pro	Thr	Glu	Ser	Val	Ser	Pro	Glu	Gly	
15		530					535					540					
	Tyr	Met	Pro	Lys	Asp	Leu	Tyr	Asn	Leu	Asn	Ser	Arg	Tyr	Gly	Thr	Ile	
	545					550					555					560	
20	Asp	Glu	Leu	Lys	Asp	Thr	Val	Lys	Lys	Phe	His	Lys	Val	Gly	Ile	Lys	
					565					570					575		
	Val	Leu	Gly	Asp	Ala	Val	Leu	Asn	His	Arg	Cys	Ala	His	Phe	Lys	Asn	
				580					585					590			
25	Gln	Asn	Gly	Val	Trp	Asn	Leu	Phe	Gly	Gly	Arg	Leu	Asn	Trp	Asp	Asp	
			595					600					605				
	Arg	Ala	Val	Val	Ala	Asp	Asp	Pro	His	Phe	Gln	Gly	Arg	Gly	Asn	Lys	
30		610					615					620					
	Ser	Ser	Gly	Asp	Asn	Phe	His	Ala	Ala	Pro	Asn	Ile	Asp	His	Ser	Gln	
	625					630					635					640	
35	Asp	Phe	Val	Arg	Lys	Asp	Ile	Lys	Glu	Trp	Leu	Cys	Trp	Met	Met	Glu	
					645					650					655		
	Glu	Val	Gly	Tyr	Asp	Gly	Trp	Arg	Leu	Asp	Phe	Val	Arg	Gly	Phe	Trp	
				660					665					670			
40	Gly	Gly	Tyr	Val	Lys	Asp	Tyr	Met	Asp	Ala	Ser	Lys	Pro	Tyr	Phe	Ala	
			675					680					685				
	Val	Gly	Glu	Tyr	Trp	Asp	Ser	Leu	Ser	Tyr	Thr	Tyr	Gly	Glu	Met	Asp	
45		690					695					700					
	Tyr	Asn	Gln	Asp	Ala	His	Arg	Gln	Arg	Ile	Val	Asp	Trp	Ile	Asn	Ala	
	705					710					715					720	
50	Thr	Ser	Gly	Ala	Ala	Gly	Ala	Phe	Asp	Val	Thr	Thr	Lys	Gly	Ile	Leu	
					725					730					735		
	His	Thr	Ala	Leu	Gln	Lys	Cys	Glu	Tyr	Trp	Arg	Leu	Ser	Asp	Pro	Lys	
				740					745					750			
55	Gly	Lys	Pro	Pro	Gly	Val	Val	Gly	Trp	Trp	Pro	Ser	Arg	Ala	Val	Thr	
			755					760					765				
	Phe	Ile	Glu	Asn	His	Asp	Thr	Gly	Ser	Thr	Gln	Gly	His	Trp	Arg	Phe	
60		770					775					780					
	Pro	Glu	Gly	Lys	Glu	Met	Gln	Gly	Tyr	Ala	Tyr	Ile	Leu	Thr	His	Pro	
	785					790					795					800	

Gly Thr Pro Ala Val Phe Phe Asp His Ile Phe Ser Asp Tyr His Ser
 805 810 815

5 Glu Ile Ala Ala Leu Leu Ser Leu Arg Asn Arg Gln Lys Leu His Cys
 820 825 830

Arg Ser Glu Val Asn Ile Asp Lys Ser Glu Arg Asp Val Tyr Ala Ala
 835 840 845

10 Ile Ile Asp Glu Lys Val Ala Met Lys Ile Gly Pro Gly His Tyr Glu
 850 855 860

Pro Pro Asn Gly Ser Gln Asn Trp Ser Val Ala Val Glu Gly Arg Asp
 865 870 875 880

15 Tyr Lys Val Trp Glu Thr Ser
 885

20 <210> 18
 <211> 1030
 <212> DNA
 <213> Malus domestica

25 <400> 18
 aggatatcag agaatgggta tgctgggctaa ggtgaatttc taagatgttc tgattttttta 60
 caagttcttc aaatttgtct cactcaataa aaacaaagat cgtacgtaaa agagcaaagc 120
 acgtactata tcctaaactt gagttatttt gcactggaaa ttatagagat tatagatttt 180
 ctgggattta catggaagtt tacctagcct tgggatttca aaatttattg tgatgagggt 240
 30 gtgttcatgc atttcttatg cacattctga cgtagaatat cagtttgtgg acgtgccctt 300
 acaaaaattc atggaccatg ttatgtgtaa tccaattga cgtgcatttc ccttttgctg 360
 caatattttt gttgaaagtt tactatgtta ctttgatata aatgtatgag tgatcgtcac 420
 agggacgata ttgggtatga cggatggagg cttgatttcg ttagaggatt ttgggggtggc 480
 tacgtcaagg actacatgga tgccagttag ccctactttg ccgtaggcga gtattgggat 540
 35 tccctaagtt atacatacgg ggaaatggat cacaatcaag atgcacacag gcagagaatt 600
 gttgattgga tcaacgctac taatggaact tgtggtgcat tcgatgtcac aacaaaaggg 660
 attctccatg cagtaagtgt gcatctttgt aatttaaaac tcaactgtta aaacttttat 720
 gaccttgctt tccatgtttg tatggcggat gatactttta ttcacatgga actacttata 780
 caggcactgg aaagatgcga gtattggcga ttgtcagata cgaagggtaa gcctccaggg 840
 40 gttcttggat ggtggccatc tcgtgctgtc actttcatag agaatcatga tactgggttct 900
 actcaggttt gaattttttt gatcttgaat tctgtcattt tacaacaag catagttata 960
 tttacaagta tcgagataca aattctttta ttcttcctt ttgaagggtc attggagggt 1020
 tccaaataaa 1030

45 <210> 19
 <211> 831
 <212> DNA
 <213> Malus domestica

50 <400> 19
 gatatcagag aatgggttatg ctgggctaagg ttaattccta acatgttctt atgcccatta 60
 tgacgtagaa tatcaatttg tggacgtacc agtaccttgc aaaaattttg gacatgttgt 120
 attttatcca attgaagtgc aatctccctt atgggtgcatg ttatgttgaa agcttatcta 180
 ttatgttgat ataaatgtat gactgaccgt cacaggaacg atattgggta cgatggatgg 240
 55 aggctggatt ttgttagagg atttcggggg ggctatgtca gggactacgt ggatgctagt 300
 gagccctact ttgctgtagg cgagtattgg gattccctaa gttatacgtg tggggaaatg 360
 gatcgcaatc aagatgcaca caggcagaga attgttgatt ggatcaacgc tactaatgga 420
 acttgtggtg catttgatgt cacaacaaaa gggattctcc atgcagtaag tgtgcatctt 480
 tgtaattttt aattgcaact gtaaaacttc tatgactttg ctttccatgt ttgtatggcg 540
 60 gatgctactt ttattcacat ggaattactt gtacttgtac aggcgctgga aagggtgcgag 600
 tattggcgat tgtcagatga gaagggtgag cctccagggg ttcttggatg gtggccatct 660
 cgtgctgcca ctttcataga gaatcatgat actggttcta ctcaggtttg aactgttttg 720
 atcatgaatt ctctgttatt ttatagtcaa gcataattaat atttaaaagt atcgagatac 780

aaattctttt attctttctt tttgaagggt cattggaggt ttccaaataa a 831

<210> 20
 <211> 667
 5 <212> DNA
 <213> Actinidia chinensis

<400> 20
 10 ccccgctcc gcttcagacc tcctcaacct ctctctgtca gagcgagctc cgccgatata 60
 gccgtggtgg agacctctga ctctgtggac gtcttattca aggagacatt cgctttgaag 120
 cgaattgaaa aggtggaggg acatatatca atcaagttag ataatgggaa agagagagaa 180
 aattggcagc tttctgtggg atgtaatctt ccggggaagt gggttcttca ctggggcgta 240
 aactatatca acgacattgg cagtgaatgg gatcagcctc ctgttgagat gaggcctcca 300
 ggctctgttc ctattaagga ttatgcaatt gaaactcctc tgaagaaatc atctgcagtg 360
 15 gtggaaggag atttatatta cgaattgaag attgatttta gtacggacaa agatattgca 420
 gctataaatt ttgttttgaa ggatgaggaa actggagcctt ggtatcagcg ccgaggaaga 480
 gatttcaaag ttgctctcat tgacgacctt catgaagatg gcaataaatt aggagctaaa 540
 aagggccttag gtgtacggcc agggcctttt gaacagctct ctagcctact gctcaaatca 600
 20 gaagaagctc atcccaaggg tgaagacaac agtgactctc gaggtcctag taaaaaaact 660
 aagtgcc 667

<210> 21
 <211> 222
 25 <212> PRT
 <213> Actinidia chinensis

<400> 21
 30 Pro Arg Val Arg Phe Arg Pro Pro Gln Pro Leu Ser Val Arg Ala Ser
 1 5 10 15
 Ser Ala Asp Thr Ala Val Val Glu Thr Ser Asp Ser Val Asp Val Leu
 20 25 30
 35 Phe Lys Glu Thr Phe Ala Leu Lys Arg Ile Glu Lys Val Glu Gly His
 35 40 45
 Ile Ser Ile Lys Leu Asp Asn Gly Lys Glu Arg Glu Asn Trp Gln Leu
 50 55 60
 40 Ser Val Gly Cys Asn Leu Pro Gly Lys Trp Val Leu His Trp Gly Val
 65 70 75 80
 Asn Tyr Ile Asn Asp Ile Gly Ser Glu Trp Asp Gln Pro Pro Val Glu
 45 85 90 95
 Met Arg Pro Pro Gly Ser Val Pro Ile Lys Asp Tyr Ala Ile Glu Thr
 100 105 110
 50 Pro Leu Lys Lys Ser Ser Ala Val Val Glu Gly Asp Leu Tyr Tyr Glu
 115 120 125
 Leu Lys Ile Asp Phe Ser Thr Asp Lys Asp Ile Ala Ala Ile Asn Phe
 130 135 140
 55 Val Leu Lys Asp Glu Glu Thr Gly Ala Trp Tyr Gln Arg Arg Gly Arg
 145 150 155 160
 Asp Phe Lys Val Ala Leu Ile Asp Asp Leu His Glu Asp Gly Asn Lys
 60 165 170 175
 Leu Gly Ala Lys Lys Gly Leu Gly Val Arg Pro Gly Pro Phe Glu Gln
 180 185 190

WO 03/027141

PCT/NZ02/00200

Leu Ser Ser Leu Leu Leu Lys Ser Glu Glu Ala His Pro Lys Gly Glu
 195 200 205

5 Asp Asn Ser Asp Ser Arg Gly Pro Ser Lys Lys Thr Lys Cys
 210 215 220

<210> 22
 10 <211> 602
 <212> DNA
 <213> Actinidia chinensis

<400> 22
 15 aaaatgggag attcctgccca agccatatcc tgctgaaaca atagttttca agaataaggc 60
 attgcggact ctattgcagc gaaaagaggg tggaaagggg gggtggagtt tatttacttt 120
 ggatgaagga tatgctggat ttgtttttgt gctcaagata aatgaaaaca catgggttgaa 180
 ttatatggga aatgactttt acatacctct ttcaagttca agtgtcttgc ctgctcaacc 240
 tagacatgat caatctgaag gtcacaggca ggtagagaca gatcaagaag tttctcctgc 300
 20 tgcataact gatggaatca tcaatgatat aagaagttta gtgagcgata tttcctctgg 360
 gaagagtcga caaacgaaat caaaagaatc tcaacaaagc attcttcaag aaattgagaa 420
 gctggctgca gaagcctaca gcactctcag aagctctatt cctacttatt cggaggatgt 480
 gatggtggaa tctgaggaag tagaaccctc tgcaaaaata tcttcaggaa caggttctgg 540
 gtttgaaatt ctctgtcaag gatttaactg ggagtcccat aaatctggaa gatggtacat 600
 25 gc 602

<210> 23
 <211> 200
 <212> PRT
 30 <213> Actinidia chinensis

<400> 23
 Lys Trp Glu Ile Pro Ala Lys Pro Tyr Pro Ala Glu Thr Ile Val Phe
 1 5 10 15

35 Lys Asn Lys Ala Leu Arg Thr Leu Leu Gln Arg Lys Glu Gly Gly Lys
 20 25 30

40 Gly Gly Trp Ser Leu Phe Thr Leu Asp Glu Gly Tyr Ala Gly Phe Val
 35 40 45

Phe Val Leu Lys Ile Asn Glu Asn Thr Trp Leu Asn Tyr Met Gly Asn
 50 55 60

45 Asp Phe Tyr Ile Pro Leu Ser Ser Ser Ser Val Leu Pro Ala Gln Pro
 65 70 75 80

Arg His Asp Gln Ser Glu Gly His Arg Gln Val Glu Thr Asp Gln Glu
 85 90 95

50 Val Ser Pro Ala Ala Tyr Thr Asp Gly Ile Ile Asn Asp Ile Arg Ser
 100 105 110

55 Leu Val Ser Asp Ile Ser Ser Gly Lys Ser Arg Gln Thr Lys Ser Lys
 115 120 125

Glu Ser Gln Gln Ser Ile Leu Gln Glu Ile Glu Lys Leu Ala Ala Glu
 130 135 140

60 Ala Tyr Ser Ile Phe Arg Ser Ser Ile Pro Thr Tyr Ser Glu Asp Val
 145 150 155 160

Met Val Glu Ser Glu Glu Val Glu Pro Pro Ala Lys Ile Ser Ser Gly

165 170 175
 Thr Gly Ser Gly Phe Glu Ile Leu Cys Gln Gly Phe Asn Trp Glu Ser
 180 185 190
 5 His Lys Ser Gly Arg Trp Tyr Met
 195 200
 10 <210> 24
 <211> 1451
 <212> DNA
 <213> Actinidia chinensis
 15 <400> 24
 ttgttgcgga tgatccacat ttccagggaa ggggcaacaa gagtagcggg gataatttcc 60
 atgctgctcc aaatattgat cattctcaag aatttgtgag gagagatctt aaagaatggc 120
 tttgttggtt aaggaaagaa attgggtatg atggatggag gcttgatttt gttcggggat 180
 tttggggagg ttatatcaag gaytacatag atgcaagtga gccttacttt gctgtaggcg 240
 20 agtattggga ttctctcagc tacacttatg gtgagatgga tcacaatcaa gatgctcata 300
 ggcagagaat tattgaatgg atcaatgcta ctagtggaac tgctggtgca tttgacgtca 360
 caactaargg aattctacat tctgcgcttc aaagatgcga gtattggcga ttatcagatc 420
 agaagggaaa acctccagga gttgttgggt ggtggccgctc tcgggcagtt acgtttatag 480
 aaaatcatga cactgggtct actcagggtc attggagatt tccagggtga aaggagatgc 540
 25 aagggtatgc atacatctg actcatcccc gaacaccagc agttttctat gatcacgctt 600
 tccatcgcat gcgatctgaa atttcggcac tcgtttcttt gagaaaccgg aacaagatcc 660
 actgtcgtag tacaattcaa ataaccaagg cagaaaggga tgtttatgca gccattatcg 720
 acaaaaaggt ggcaatgaag ataggcccag gtttctacga acccgcaagt gggcccaaaa 780
 gatggtctct ggccgttgag ggaaacgatt acaaagtctg ggaagcgtca tagaatttga 840
 30 aaatgcattg tcttgaagca catgaagttt atagaacatc atttcgacca aaggagagat 900
 acaaaatgat attatctata ataaagttga acccttcccc caacggcata cttcaattga 960
 agaccgcatt gcagcacatg ttcaccttgt ttgaggaagt gatcatgtcc acgtgatgat 1020
 atgtgggaat tgattggggg ttaagtagaa atatacataa tacatacttt tgtaatctta 1080
 cacttatatt gaattaccag tgtaaactct tgaacatctg gaaaataaat tcccaagagg 1140
 35 caagagggta ttgtaagaac tcggacagac agtgtaaaac tgttctataa atttgtagca 1200
 tagtccaaat aagctcatta tattataaag gtgagttatc tacaaacat tctccattga 1260
 ttattgtgtt tgcacagtga aacgagatga tatggaaatc tttgaaacta aactgagaa 1320
 gtctgatgtc gactgaatcc tctgttgtga tttctgcttt gtacgtggag tttttaaggc 1380
 ttcttaaaag gattggattg tcattcaatt aagtgcgga aagagactat gtgacctcat 1440
 40 ggaaagaagc c 1451
 <210> 25
 <211> 276
 <212> PRT
 45 <213> Actinidia chinensis
 <400> 25
 Val Ala Asp Asp Pro His Phe Gln Gly Arg Gly Asn Lys Ser Ser Gly
 1 5 10 15
 50 Asp Asn Phe His Ala Ala Pro Asn Ile Asp His Ser Gln Glu Phe Val
 20 25 30
 Arg Arg Asp Leu Lys Glu Trp Leu Cys Trp Leu Arg Lys Glu Ile Gly
 35 40 45
 55 Tyr Asp Gly Trp Arg Leu Asp Phe Val Arg Gly Phe Trp Gly Gly Tyr
 50 55 60
 60 Ile Lys Asp Tyr Ile Asp Ala Ser Glu Pro Tyr Phe Ala Val Gly Glu
 65 70 75 80
 Tyr Trp Asp Ser Leu Ser Tyr Thr Tyr Gly Glu Met Asp His Asn Gln

	85	90	95
5	Asp Ala His Arg Gln Arg Ile Ile Glu Trp Ile Asn Ala Thr Ser Gly 100 105 110		
	Thr Ala Gly Ala Phe Asp Val Thr Thr Lys Gly Ile Leu His Ser Ala 115 120 125		
10	Leu Gln Arg Cys Glu Tyr Trp Arg Leu Ser Asp Gln Lys Gly Lys Pro 130 135 140		
	Pro Gly Val Val Gly Trp Trp Pro Ser Arg Ala Val Thr Phe Ile Glu 145 150 155 160		
15	Asn His Asp Thr Gly Ser Thr Gln Gly His Trp Arg Phe Pro Gly Gly 165 170 175		
	Lys Glu Met Gln Gly Tyr Ala Tyr Ile Leu Thr His Pro Gly Thr Pro 180 185 190		
20	Ala Val Phe Tyr Asp His Ala Phe His Arg Met Arg Ser Glu Ile Ser 195 200 205		
	Ala Leu Val Ser Leu Arg Asn Arg Asn Lys Ile His Cys Arg Ser Thr 210 215 220		
25	Ile Gln Ile Thr Lys Ala Glu Arg Asp Val Tyr Ala Ala Ile Ile Asp 225 230 235 240		
30	Lys Lys Val Ala Met Lys Ile Gly Pro Gly Phe Tyr Glu Pro Ala Ser 245 250 255		
	Gly Pro Gln Arg Trp Ser Leu Ala Val Glu Gly Asn Asp Tyr Lys Val 260 265 270		
35	Trp Glu Ala Ser 275		
40	<210> 26 <211> 450 <212> DNA <213> Actinidia chinensis		
45	<400> 26 ggcttctctg gtgtatatgt gaaagaatat atagaagctt caaatcctgc ttttgcaatt 60 ggagaatatt gggatagttt ggcttacgag ggtggtgatt tgtgttacaa ccaagatgct 120 catcgacaac ggattgttaa ttggatcaat gccactggcg gtgcctcatc agcatttgac 180 gtcacaacaa agggatatatt acattctgct ctgcataatc aatattggag gttgattgat 240 50 cctcaaggaa aaccaacagg ggtaatgggg tgggtggccat cgcgtgctgt cacattctta 300 gagaaccatg atacaggatc tacacagggt cattggccat ttccaagaga caagcttgca 360 cagggatacg cttacatttt aaccatccg ggaacgcctg taatatttta tgaccatttc 420 tatgactttg ggctccacga cgtcataacc 450		
55	<210> 27 <211> 150 <212> PRT <213> Actinidia chinensis		
60	<400> 27 Gly Phe Ser Gly Val Tyr Val Lys Glu Tyr Ile Glu Ala Ser Asn Pro 1 5 10 15		

Ala Phe Ala Ile Gly Glu Tyr Trp Asp Ser Leu Ala Tyr Glu Gly Gly
20 25 30

5 Asp Leu Cys Tyr Asn Gln Asp Ala His Arg Gln Arg Ile Val Asn Trp
35 40 45

Ile Asn Ala Thr Gly Gly Ala Ser Ser Ala Phe Asp Val Thr Thr Lys
50 55 60

10 Gly Ile Leu His Ser Ala Leu His Asn Gln Tyr Trp Arg Leu Ile Asp
65 70 75 80

Pro Gln Gly Lys Pro Thr Gly Val Met Gly Trp Trp Pro Ser Arg Ala
85 90 95

15 Val Thr Phe Leu Glu Asn His Asp Thr Gly Ser Thr Gln Gly His Trp
100 105 110

20 Pro Phe Pro Arg Asp Lys Leu Ala Gln Gly Tyr Ala Tyr Ile Leu Thr
115 120 125

His Pro Gly Thr Pro Val Ile Phe Tyr Asp His Phe Tyr Asp Phe Gly
130 135 140

25 Leu His Asp Val Ile Thr
145 150

30 <210> 28
<211> 592
<212> DNA
<213> Vaccinium corymbosum

35 <400> 28
tccgagggcat taaaacagcc cccacaaata agttcgggaa cagggttcagg ttttgaaatt 60
ctgtgccaaag gattttaactg ggagtctcat aaatccggaa gatggtacat ggagctcaaa 120
gaaaaaattg ctgaaatgtc ttctcttggt ttactgtgg tttgggtacc accaccgaca 180
gattctgtct cgcctgaagg ctacatgcca agggatttgt atgacttgaa ctccagatac 240
gggtggtactg atgagctgaa ggttctgggt aagagattcc atgaagttgg tatcaaagtt 300
40 cttggagatg ttgtcctaaa tcaccgttgc gcacaatata agaatcaaaa tgggtgtatgg 360
aatattttcg gtggtcgtct aaattgggat gatcgtgctg ttgttgcaga tgatccacat 420
ttccagggaa gaggcaacaa gagtagtgga gataatttcc atgctgcacc aaatatagat 480
cattcgcagg aatttgtgag gaaggatctt aaaaaatggc tacgctggct aaggaaaaat 540
attgggtatg atggatggag gcttgatttc gctcggggat tctggggggg tt 592

45 <210> 29
<211> 197
<212> PRT
<213> Vaccinium corymbosum

50 <400> 29
Ser Glu Ala Leu Lys Gln Pro Pro Gln Ile Ser Ser Gly Thr Gly Ser
1 5 10 15

55 Gly Phe Glu Ile Leu Cys Gln Gly Phe Asn Trp Glu Ser His Lys Ser
20 25 30

Gly Arg Trp Tyr Met Glu Leu Lys Glu Lys Ile Ala Glu Met Ser Ser
35 40 45

60 Leu Gly Phe Thr Val Val Trp Leu Pro Pro Pro Thr Asp Ser Val Ser
50 55 60

Pro Glu Gly Tyr Met Pro Arg Asp Leu Tyr Asp Leu Asn Ser Arg Tyr
65 70 75 80

5 Gly Gly Thr Asp Glu Leu Lys Val Leu Val Lys Arg Phe His Glu Val
85 90 95

Gly Ile Lys Val Leu Gly Asp Val Val Leu Asn His Arg Cys Ala Gln
100 105 110

10 Tyr Lys Asn Gln Asn Gly Val Trp Asn Ile Phe Gly Gly Arg Leu Asn
115 120 125

Trp Asp Asp Arg Ala Val Val Ala Asp Asp Pro His Phe Gln Gly Arg
130 135 140

15 Gly Asn Lys Ser Ser Gly Asp Asn Phe His Ala Ala Pro Asn Ile Asp
145 150 155 160

His Ser Gln Glu Phe Val Arg Lys Asp Leu Lys Lys Trp Leu Arg Trp
20 165 170 175

Leu Arg Lys Asn Ile Gly Tyr Asp Gly Trp Arg Leu Asp Phe Ala Arg
180 185 190

25 Gly Phe Trp Gly Gly
195

30 <210> 30
<211> 963
<212> DNA
<213> Coffea arabica

35 <400> 30
atattatgcc aaggttttaa ctgggagtca cataaatctg gaagatggta tatggagctt 60
catcaaaaag ctgctgagtt atcatcactt gggttttactg ttgtctgggtt acctccgccc 120
acagagtctg tttcacctga aggctacatg ccaaaggatt tgtacaatct gaactccagg 180
taagaacagt tgtgttgga aactaatata ggtactaaga ttagattagt ttcctggaat 240
cctgaatttg gacatgctta tgtagttttc ctgttcagat aatttaccat atgtagaatg 300
40 taggactgaa ctctgacttt agacctaat gtgcaaactt cattaatact gtttgcacct 360
acttcaatcc tatgaactaa cataatcatg aaaaccatgg ccttcttaat agtacaatct 420
gctgtagtct tcttcatgga tttgagttga aaatgggtag ctaaataaaa tctattgggg 480
tagctgacac ctgtctactt tcagaagtag gctattcatt cttcttttca tgcctgtctc 540
ccaattcttt atgctactgg gcagatgtta attaaatttt ttatattctc ttgtaatttc 600
45 ttttttatct ctaggtaaag ttataatttg tttatttgat tagaataact aaactaaggg 660
aaciaataaa tctgattact tgaaattgct ctgattattg gaataaatgc tgatcaacga 720
ataactactg aaattgttat accacatctt cttgcagcac tctccctctg tctgtctgcc 780
tctgtgggtg atgtttttta cttcttgatc attttcagat atggaagtat tgatgaattg 840
aagtctctcg tgaagagatt ccatgaagtt ggcacatcag tccctcgaga tgctgtgcta 900
50 aaccatagat gtgcccatta caaaaatcag aatgggtattt ggaatatatt tggaggtcgc 960
cta 963

55 <210> 31
<211> 108
<212> PRT
<213> Coffea arabica

60 <400> 31
Ile Leu Cys Gln Gly Phe Asn Trp Glu Ser His Lys Ser Gly Arg Trp
1 5 10 15

Tyr Met Glu Leu His Gln Lys Ala Ala Glu Leu Ser Ser Leu Gly Phe
20 25 30

WO 03/027141

PCT/NZ02/00200

Thr Val Val Trp Leu Pro Pro Pro Thr Glu Ser Val Ser Pro Glu Gly
 35 40 45

5 Tyr Met Pro Lys Asp Leu Tyr Asn Leu Asn Ser Arg Tyr Gly Ser Ile
 50 55 60

Asp Glu Leu Lys Ser Leu Val Lys Arg Phe His Glu Val Gly Ile Met
 65 70 75 80

10 Val Leu Gly Asp Ala Val Leu Asn His Arg Cys Ala His Tyr Lys Asn
 85 90 95

15 Gln Asn Gly Ile Trp Asn Ile Phe Gly Gly Arg Leu
 100 105

<210> 32
 <211> 810
 20 <212> DNA
 <213> Gossypium hirsutum

<400> 32
 25 atattgtgcc aaggttttaa ctgggaatct aataaatctg gacgatggta tatggaactc 60
 aaagaaaaag cttctgaaat atcttcactt gggtttactg taatttggct gccaccacca 120
 accgagtctg tgtcacctga aggttacatg ccgaaggact tgtataattt aaattccagg 180
 tacatattgc aagcttgtag ttacataac aaatctttca ggtttttaca ttttggtgcc 240
 tgggtcaaatt caaagaagct gaaacttgca aatatgaact ttcaaagtct ttttttgatt 300
 cttttttcta aaaattataa ttgggaaact ttgttcaggg aaaaagatgg aactgaatgt 360
 30 tttttatggt tgcttttagta tgctgatcaa tgggttggtg caagtctcag attttggttc 420
 aacccaaaaga aagtgcacat ttttttacct ctcttttatg cttctttaac ttacatttaa 480
 ccactctgtaa tagtaataa agacttcaat tactctttta atctaaaaga ctacatgcaa 540
 atgcctcttg gacatgctga aatgtagaaa gtcattggtta tttcccatat ctgaaagaaa 600
 tcatgttttc atgctcctca agaggcctaa gcaaaccagt gtgcaactgt tttttcactg 660
 35 atcaggtatg gaacaatcga tgagctgaag gacctcgtaa aaagcctcca tggagttggt 720
 ttgaaagttc ttggtgatgt tgtattgaat caccgttggt cacactatca aaatcagaat 780
 ggtgtttgga atctatttgg tggacgccta 810

<210> 33
 40 <211> 108
 <212> PRT
 <213> Gossypium hirsutum

<400> 33
 45 Ile Leu Cys Gln Gly Phe Asn Trp Glu Ser Asn Lys Ser Gly Arg Trp
 1 5 10 15

Tyr Met Glu Leu Lys Glu Lys Ala Ser Glu Ile Ser Ser Leu Gly Phe
 20 25 30

50 Thr Val Ile Trp Leu Pro Pro Pro Thr Glu Ser Val Ser Pro Glu Gly
 35 40 45

55 Tyr Met Pro Lys Asp Leu Tyr Asn Leu Asn Ser Arg Tyr Gly Thr Ile
 50 55 60

Asp Glu Leu Lys Asp Leu Val Lys Ser Leu His Gly Val Gly Leu Lys
 65 70 75 80

60 Val Leu Gly Asp Val Val Leu Asn His Arg Cys Ala His Tyr Gln Asn
 85 90 95

Gln Asn Gly Val Trp Asn Leu Phe Gly Gly Arg Leu

100

105

5 <210> 34
 <211> 600
 <212> DNA
 <213> Rosa woodsii

10 <400> 34
 cctatgataa atttatacta caaaataatc ttgcatgata ttcataagctt ttactgaccc 60
 caaataaata agctagctct acactgcatt tttttgcagt aactttatac agaaaaaggc 120
 caacttttgt tgggtaaaaa aaaaaaatta aggtgggttc gcaaacctct gcttgctgg 180
 ttatgcactt ctaatcagtt ggaccaagca tgtactgacc atcaaagttt cagaccttat 240
 ccttcttttt ccttttggtc tgaactctga agtcttgaag atgaaccctt cttcttcttc 300
 15 actctctctc tttttttttt ggtttacctt ggttggtagt aaattccact cctttgtttc 360
 gggagttcca tttcttttagt attcttaata ttgttctggg agttccggag ggatgctgta 420
 tacatcagtg attacttttg tttcacactt ctcagatatg gcactatgga tgaactgaag 480
 gagactgtga aggcattcca taaagttggg atcaaagtac ttggagatgt tgttctaaat 540
 caccgctgtg cgcagtatca gaatagcaat ggtgtatgga atctatttgg aggacgcctt 600

20 <210> 35
 <211> 48
 <212> PRT
 <213> Rosa woodsii

25 <400> 35
 Tyr Gly Thr Met Asp Glu Leu Lys Glu Thr Val Lys Ala Phe His Lys
 1 5 10 15

30 Val Gly Ile Lys Val Leu Gly Asp Val Val Leu Asn His Arg Cys Ala
 20 25 30

Gln Tyr Gln Asn Ser Asn Gly Val Trp Asn Leu Phe Gly Gly Arg Leu
 35 40 45

40 <210> 36
 <211> 22
 <212> DNA
 <213> Artificial Sequence

45 <220>
 <223> Description of Artificial Sequence: Made in lab

50 <400> 36
 aacggtgatt cagaacagca tc 22

55 <210> 37
 <211> 22
 <212> DNA
 <213> Artificial Sequence

60 <220>
 <223> Description of Artificial Sequence: Made in lab

<400> 37
 atgtaccctt gaggcgacac ag 22

<210> 38
 <211> 21

<212> DNA
 <213> Artificial Sequence

5 <220>
 <223> Description of Artificial Sequence: Made in lab

<400> 38
 ggtgggaacc aaatcacagt g 21

10 <210> 39
 <211> 35
 <212> DNA
 <213> Artificial Sequence

15 <220>
 <223> Description of Artificial Sequence: Made in lab

<400> 39
 gactcgagtc gacatcgatt tttttttttt ttttt 35

20 <210> 40
 <211> 17
 <212> DNA
 <213> Artificial Sequence

25 <220>
 <223> Description of Artificial Sequence: Made in lab

<400> 40
 30 gactcgagtc gacatcg 17

<210> 41
 <211> 26
 35 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Made in lab

40 <400> 41
 atatrttgcc aaggttttaa ctggga 26

<210> 42
 45 <211> 24
 <212> PRT
 <213> Artificial Sequence

<220>
 50 <223> Description of Artificial Sequence: Made in lab

<400> 42
 Trp Ala Arg Arg Cys Gly Trp Cys Cys Trp Cys Cys Ala Ala Ala Trp
 1 5 10 15

55 Ala Lys Ala Thr Thr Cys Cys Ala
 20

60 <210> 43
 <211> 23
 <212> DNA
 <213> Artificial Sequence

WO 03/027141

PCT/NZ02/00200

<220>
 <223> Description of Artificial Sequence: Made in lab

5 <400> 43
 gttacttacc gggaaagcta tac 23

<210> 44
 <211> 18
 10 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Made in lab

15 <400> 44
 gggcgctcca tcaaaatc 18

<210> 45
 20 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 25 <223> Description of Artificial Sequence: Made in lab

<400> 45
 tcttccacag gacctttatt c 21

30 <210> 46
 <211> 21
 <212> DNA
 <213> Artificial Sequence

35 <220>
 <223> Description of Artificial Sequence: Made in lab

<400> 46
 agtcctccgg tacaagaagt c 21

40 <210> 47
 <211> 26
 <212> DNA
 <213> Artificial Sequence

45 <220>
 <223> Description of Artificial Sequence: Made in lab

<400> 47
 50 ggaacaattg atgagctaaa agatac 26

<210> 48
 <211> 19
 <212> DNA
 55 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Made in lab

60 <400> 48
 ggaaatgagg gtcattctgc 19

<210> 49

WO 03/027141

PCT/NZ02/00200

<211> 22
 <212> DNA
 <213> Artificial Sequence

5 <220>
 <223> Description of Artificial Sequence: Made in lab

<400> 49
 aacggtgatt cagaacagca tc 22

10 <210> 50
 <211> 22
 <212> DNA
 <213> Artificial Sequence

15 <220>
 <223> Description of Artificial Sequence: Made in lab

<400> 50
 20 atgtaccctt gaggcgacac ag 22

<210> 51
 <211> 21
 <212> DNA
 25 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Made in lab

30 <400> 51
 ggtgggaacc aaatcacagt g 21

<210> 52
 <211> 2622
 35 <212> DNA
 <213> Oryza sativa

<400> 52
 40 atggcgggtgg cgagctggag cattccggcg atcccgcggg cgggtccgac agcgaggggt 60
 gtgctgctcg gcggcgccctt cgtgacggcc gcgcggccgc ccgtggcgtg gcggtgccgg 120
 gccacgctcc ctaggagagt gaggtcgggt ggcgtgggtg cccgtgccgg tgcggcggag 180
 acgccgggtg ccggctccgg ggaagccggg ttgctgttct ccgagaagtt ccccttgccg 240
 cgatcccgaa cgggtggaagg gaaggcgtgg gtgaggggtcg atgcggagcc ggatggggag 300
 ggcaagtgca aggttgtgat cgggtgtgat gtagagggga agtgggtgct gcattgggggt 360
 45 gtctcctacg atggtgagca ggaagagaa tgggaccaac ctcttcaga catgagacct 420
 cccggttcag tgcctattaa ggactatgca attgaaacat ctttggacac tccacacaat 480
 tcagaaggca agacgattca tgaagtgcaa atcaaaattg ataagggcac atcaattgct 540
 gctatcaatt ttgttctaaa ggaagaggaa acaggtgctt ggtttcagca caaggggtcag 600
 gatttcagaa tacctttaag tggatccctt ggtggagatc tactaggaac agaacaagat 660
 50 attgatgtca ggccagggtca cctatctaac gtgttacaga aacctgaggg acctattgct 720
 gagcctcata aaactgtacc cgatgataaa ggttcaagaa ccaaacacat ttcagggtttc 780
 tatgaggaat acccaatctt aaaaacgggtg tatgttcaga attttataac tgttaatgtg 840
 agggaaaaca atggaacaac taaacatgct gtggaattcg aactgatata tcctggagaa 900
 gttatcattc attggggagt ttgcaaagac aataccatga catgggagat ccccccagaa 960
 55 ccacatccac ctgcaacgaa gatattccga cagaaagctc ttcagaccat gctacaacaa 1020
 aaagctgatg gaacaggcaa ctctctatca ttcttactgg atggagagta ttctgggtctg 1080
 atttttgtgg taaaacttga tgagtatact tgggttgagaa atgtggagaa tggatttgat 1140
 ttctacattc ctcttacaag agcagacgcc gaggtgaca aacagaaagc cgatgataag 1200
 tcttcacaag atgatggctt aatcagtgat ataaggaatc tgggtggttg gctgtcgtct 1260
 60 agaagagggtc agcgagcgaa gaataaagtt ctgcaagagg atatcctaca agaaattgag 1320
 aggttagcgg cagaagctta tagcattttt aggagcccca caattgatac tgtagaggaa 1380
 tctgtttaca ttgatgactc atccattgtg aagcctgctt gttctggtac tggatctgga 1440
 tttgaaatat tgtgtcaagg atttaactgg gaatctcata agtcaggaaa atggtatgtg 1500

```

gaacttggct caaaggccaa ggagttgtca tcatggggtt tcaccattgt ctggtcacca 1560
ccacctactg attctgtgtc gcctgaagga tacatgccaa gggatttgta taatctaaat 1620
tccagatatg ggaccatgga agagttgaag gaggctgtga aacgttttca tgaagccggt 1680
atgaagggtc ttggtgatgc cgtcctgaat cacaggtgtg ctcaatttca gaacccaaaat 1740
5  ggcgctctgga atatttttgg tggacgcctt aactgggatg atcgagcagt tgttgcagat 1800
gatccacatt tccaggggaag aggaaacaag agcagtggag ataacttcca tgcagcccca 1860
aacattgatc actcgcaaga gtttgtgagg agtgatctta aagaatggct ttgttggatg 1920
agaaaggaag ttggatacga tggatggcga cttgattttg ttcgcggatt ttgggggtgga 1980
tatgtccacg attacttgga agcaagcgaa ccatattttg cagtaggaga gtactgggat 2040
10  tctctcagtt acacctatgg tgaaatggat tataatcaag atgccacag gcagagaata 2100
gttgattgga taaatgctac aaatgggaact gctgggtgcat ttgatgttac cacgaaagga 2160
atacttcact ctgcactgga aagatctgag tactggcgct tgtctgatga aaaaggaaaa 2220
ccccctggag tgttagggtg gtggccttcg cgtgctgtca catttataga aaatcatgac 2280
actggttcta ctcaggggtc ttggagattc ccctttggta tggagttgca aggctatgtc 2340
15  tacatcttaa ctcaccagc cactcctgca atcttctatg atcatatatt ttcgcattta 2400
cagccagaga ttgctaaatt aatttctatt agaatcgcc aaaagatcca ttgccgtagc 2460
aagatcaaga tactgaaagc agagggaaat ttatatgcgg cagagattga tgagagggta 2520
acaatgaaga ttggcgcagg acattttgag ccaagcggcc ccacaaactg ggtagttgct 2580
gccgagggac aggattacaa ggtctgggaa gtgtcatcgt ag 2622
20
<210> 53
<211> 873
<212> PRT
<213> Oryza sativa
25
<400> 53
Met Ala Val Ala Ser Trp Ser Ile Pro Ala Ile Pro Arg Ala Gly Pro
1 5 10 15
30  Thr Ala Arg Gly Val Leu Leu Gly Gly Ala Phe Val Thr Ala Ala Arg
20 25 30
Pro Pro Val Ala Trp Arg Cys Arg Ala Thr Leu Pro Arg Arg Val Arg
35 35 40 45
Leu Gly Gly Val Val Ala Arg Ala Gly Ala Ala Glu Thr Pro Val Ala
50 55 60
40  Gly Ser Gly Glu Ala Gly Leu Leu Phe Ser Glu Lys Phe Pro Leu Arg
65 70 75 80
Arg Ser Arg Thr Val Glu Gly Lys Ala Trp Val Arg Val Asp Ala Glu
85 90 95
45  Pro Asp Gly Glu Gly Lys Cys Lys Val Val Ile Gly Cys Asp Val Glu
100 105 110
Gly Lys Trp Val Leu His Trp Gly Val Ser Tyr Asp Gly Glu Gln Gly
115 120 125
50  Arg Glu Trp Asp Gln Pro Pro Ser Asp Met Arg Pro Pro Gly Ser Val
130 135 140
Pro Ile Lys Asp Tyr Ala Ile Glu Thr Ser Leu Asp Thr Pro His Asn
145 150 155 160
55  Ser Glu Gly Lys Thr Ile His Glu Val Gln Ile Lys Ile Asp Lys Gly
165 170 175
60  Thr Ser Ile Ala Ala Ile Asn Phe Val Leu Lys Glu Glu Glu Thr Gly
180 185 190
Ala Trp Phe Gln His Lys Gly Gln Asp Phe Arg Ile Pro Leu Ser Gly

```

	195	200	205
	Ser Phe Gly Gly Asp Leu Leu Gly Thr Glu Gln Asp Ile Asp Val Arg		
5	210	215	220
	Pro Gly His Leu Ser Asn Val Leu Gln Lys Pro Glu Gly Pro Ile Ala		
	225	230	235
10	Glu Pro His Lys Thr Val Pro Asp Asp Lys Gly Ser Arg Thr Lys His		
	245	250	255
	Ile Ser Gly Phe Tyr Glu Glu Tyr Pro Ile Leu Lys Thr Val Tyr Val		
	260	265	270
15	Gln Asn Phe Ile Thr Val Asn Val Arg Glu Asn Asn Gly Thr Thr Lys		
	275	280	285
	His Ala Val Glu Phe Asp Thr Asp Ile Pro Gly Glu Val Ile Ile His		
20	290	295	300
	Trp Gly Val Cys Lys Asp Asn Thr Met Thr Trp Glu Ile Pro Pro Glu		
	305	310	315
25	Pro His Pro Pro Ala Thr Lys Ile Phe Arg Gln Lys Ala Leu Gln Thr		
	325	330	335
	Met Leu Gln Gln Lys Ala Asp Gly Thr Gly Asn Ser Leu Ser Phe Leu		
	340	345	350
30	Leu Asp Gly Glu Tyr Ser Gly Leu Ile Phe Val Val Lys Leu Asp Glu		
	355	360	365
	Tyr Thr Trp Leu Arg Asn Val Glu Asn Gly Phe Asp Phe Tyr Ile Pro		
35	370	375	380
	Leu Thr Arg Ala Asp Ala Glu Ala Asp Lys Gln Lys Ala Asp Asp Lys		
	385	390	395
40	Ser Ser Gln Asp Asp Gly Leu Ile Ser Asp Ile Arg Asn Leu Val Val		
	405	410	415
	Gly Leu Ser Ser Arg Arg Gly Gln Arg Ala Lys Asn Lys Val Leu Gln		
	420	425	430
45	Glu Asp Ile Leu Gln Glu Ile Glu Arg Leu Ala Ala Glu Ala Tyr Ser		
	435	440	445
	Ile Phe Arg Ser Pro Thr Ile Asp Thr Val Glu Glu Ser Val Tyr Ile		
50	450	455	460
	Asp Asp Ser Ser Ile Val Lys Pro Ala Cys Ser Gly Thr Gly Ser Gly		
	465	470	475
55	Phe Glu Ile Leu Cys Gln Gly Phe Asn Trp Glu Ser His Lys Ser Gly		
	485	490	495
	Lys Trp Tyr Val Glu Leu Gly Ser Lys Ala Lys Glu Leu Ser Ser Met		
	500	505	510
60	Gly Phe Thr Ile Val Trp Ser Pro Pro Pro Thr Asp Ser Val Ser Pro		
	515	520	525
	Glu Gly Tyr Met Pro Arg Asp Leu Tyr Asn Leu Asn Ser Arg Tyr Gly		

	530		535		540											
	Thr	Met	Glu	Glu	Leu	Lys	Glu	Ala	Val	Lys	Arg	Phe	His	Glu	Ala	Gly
	545					550					555					560
5	Met	Lys	Val	Leu	Gly	Asp	Ala	Val	Leu	Asn	His	Arg	Cys	Ala	Gln	Phe
					565					570					575	
	Gln	Asn	Gln	Asn	Gly	Val	Trp	Asn	Ile	Phe	Gly	Gly	Arg	Leu	Asn	Trp
10				580					585					590		
	Asp	Asp	Arg	Ala	Val	Val	Ala	Asp	Asp	Pro	His	Phe	Gln	Gly	Arg	Gly
			595					600					605			
15	Asn	Lys	Ser	Ser	Gly	Asp	Asn	Phe	His	Ala	Ala	Pro	Asn	Ile	Asp	His
	610						615					620				
	Ser	Gln	Glu	Phe	Val	Arg	Ser	Asp	Leu	Lys	Glu	Trp	Leu	Cys	Trp	Met
	625					630					635					640
20	Arg	Lys	Glu	Val	Gly	Tyr	Asp	Gly	Trp	Arg	Leu	Asp	Phe	Val	Arg	Gly
					645					650					655	
	Phe	Trp	Gly	Gly	Tyr	Val	His	Asp	Tyr	Leu	Glu	Ala	Ser	Glu	Pro	Tyr
25				660					665					670		
	Phe	Ala	Val	Gly	Glu	Tyr	Trp	Asp	Ser	Leu	Ser	Tyr	Thr	Tyr	Gly	Glu
			675					680					685			
30	Met	Asp	Tyr	Asn	Gln	Asp	Ala	His	Arg	Gln	Arg	Ile	Val	Asp	Trp	Ile
	690						695					700				
	Asn	Ala	Thr	Asn	Gly	Thr	Ala	Gly	Ala	Phe	Asp	Val	Thr	Thr	Lys	Gly
	705					710					715					720
35	Ile	Leu	His	Ser	Ala	Leu	Glu	Arg	Ser	Glu	Tyr	Trp	Arg	Leu	Ser	Asp
					725					730					735	
	Glu	Lys	Gly	Lys	Pro	Pro	Gly	Val	Leu	Gly	Trp	Trp	Pro	Ser	Arg	Ala
40				740					745					750		
	Val	Thr	Phe	Ile	Glu	Asn	His	Asp	Thr	Gly	Ser	Thr	Gln	Gly	His	Trp
			755					760					765			
45	Arg	Phe	Pro	Phe	Gly	Met	Glu	Leu	Gln	Gly	Tyr	Val	Tyr	Ile	Leu	Thr
	770						775					780				
	His	Pro	Gly	Thr	Pro	Ala	Ile	Phe	Tyr	Asp	His	Ile	Phe	Ser	His	Leu
	785					790				795						800
50	Gln	Pro	Glu	Ile	Ala	Lys	Leu	Ile	Ser	Ile	Arg	Asn	Arg	Gln	Lys	Ile
					805					810					815	
	His	Cys	Arg	Ser	Lys	Ile	Lys	Ile	Leu	Lys	Ala	Glu	Gly	Asn	Leu	Tyr
55				820					825					830		
	Ala	Ala	Glu	Ile	Asp	Glu	Arg	Val	Thr	Met	Lys	Ile	Gly	Ala	Gly	His
			835					840					845			
60	Phe	Glu	Pro	Ser	Gly	Pro	Thr	Asn	Trp	Val	Val	Ala	Ala	Glu	Gly	Gln
	850						855					860				
	Asp	Tyr	Lys	Val	Trp	Glu	Val	Ser	Ser							

865

870

5 <210> 54
 <211> 8602
 <212> DNA
 <213> *Oryza sativa*

10 <400> 54
 atggcggtgg cgagctggag cattccggcg atccccgagg cgggtccgac agcgaggggt 60
 gtgctgctcg gcggcgccctt cgtgacggcc gcgcggccgc ccgtggcgtg gcggtgccgg 120
 gccacgctcc ctaggagagt gaggtcgggt ggcgtgggtg cccgtgccgg tgcggcggag 180
 acgccggtgg ccggctccgg ggaagccggg ttgctgttct ccgagaagtt ccccttgccg 240
 cgatcccgaa cggtaacttg ttctgctctg ttgttgat ctgcggcgaa gctatgcttc 300
 15 gttgtttttt tttgttggt gatggggtct gtggtgttca ggtggaaggg aaggcgtggg 360
 tgagggtcga tgcggagccg gatggggagg gcaagtgcaa ggttgatgc ggggtgatg 420
 tagaggggaa gtgggtgctg cattgggggtg tctcctacga tggtagcag ggaaggact 480
 cgctgattct ctctccgagc tgtgtctctg tttgcatgat ttatgaacta tgagtgcac 540
 cttgttttga tcatccgtgg ggagttgatt tgtgcattgt tgagagcgtc gtttgaatca 600
 20 acgatgggtg tcagactggc aggtatgtcc tcatggtagg ctgctttcca tatctatgta 660
 gcctcacccg tccagtctaa gcacttgat atgtatatag tgaaattact gtcaagtgtc 720
 aatgcaaccc cttaatcttt atttggttc gattagagga aattataaac aggttcatta 780
 tgaaaaattc ttacaagggtg ggtgccattg atttatggct atacttgga aatgcgcgcg 840
 aaaagagaat tttgattagg tgacggggta gtgccgtaac catgtaggtt ggctcttata 900
 25 ctgtttttct tttttatttg gaatttcgat atagcgttga agcgcgaggt gattagaaat 960
 aggcatactt gcaagatagg aatgataagt tcttaaattt caattgcagg ttgcccggtg 1020
 ttatgtctcc ttatctcctt tggagttgtc ttgatattcc aatattctat acgttactta 1080
 aatggcacga tagttttgca aggaatatct agaattaata tttgtcatat cttttgtttt 1140
 tgtatgtatt caagcttcaa catacataca tttctccccg caaaaaaaaaa agaacttttc 1200
 30 tgtaaaaggc aaaaaattga attttgtcgt ctttcagatt tagttgttgc tgtaatgtcc 1260
 ttcccttgaat gcaaaaataa atgatctagc tcctgaaata tatggcattc tttatcctta 1320
 ttttagtgacc caatttattg ttatcacttg tttctataat tttttgcaga gaatgggacc 1380
 aacctccttc agacatgaga cctcccggtt cagtgcctat taaggttgtt ctaatcctat 1440
 gaaaaacatg ttaagaact cactactgac attcacatcc caggatctc taagtttgtt 1500
 35 tgatgtctga gtcgaaagct cgttgcatcc tgtagccata ttttttctca tttgatcttc 1560
 ctattgctga acaatatctc aaactttcag gactatgcaa ttgaaacatc tttggacact 1620
 ccacacaatt cagaaggcaa gacgattcat gaagtgcata tcaaaattga taagggcaca 1680
 tcaattgctg ctatcaattt tgttctaaag gttcaaattt tacgctgctg tatgctttct 1740
 ctgctgtcag ttttatttcc taggcattct gtaccatgtg tcaaaaggcc tagaagtcta 1800
 40 cgactggccc attagattcg taaaattgct gaaagtgcct aagggtggga gttgtcttcc 1860
 catttttttag ttgtctaaga aaagaataat atataacccc tacagggtata tctgcatact 1920
 aaagtattgg gagcggatat actaaacaca tgcctcttag atctatagat tagttgttaa 1980
 atgcactact ctttctttta attactacat ttttgcatc atcttttgat ttgaatgtct 2040
 cgatggactg ctgctttgct aggaagagga aacagggtgct tggtttcagc acaagggtca 2100
 45 ggatttcaga atacctttta gtggatcctt tgggtggagt ctactaggaa cagaacaaga 2160
 tattgatgtc aggccagggt aggaagtatt ttagtgaatc atgtttttaga ttattactat 2220
 agtggtttgg cactttgttc ttttcattgg ttggttttta tgttatataa aaatgcattt 2280
 ccttctataa gaaaataatc ttatttaggg ttaaaatctt agtaattttc ataggatgct 2340
 tgatcatgaa tattaccaca ttgaaaattt accccatttg cacttttcat atttcagtta 2400
 50 tctaacatac tcacatctct gatttcttag gggctttagg tcacctatct aacgtgttac 2460
 agaaacctga gggacctatt gctgagcctc ataaaactgt acccgatgat aaaggttcaa 2520
 gaaccaaaca catttcagggt ttctatgagg aatacccaat cttaaaaacg gtgtatgttc 2580
 agaattttat aactgttaat gtgagggaaa acaatggaac aactaaacat gctgtggaat 2640
 tcgacactga tattcctgga gaagtattca ttcatgggg agtttgcaaa gacaatacca 2700
 55 tgacatggga gatccccca gaaccacatc cacctgcaac gaagatattc cgacagaaag 2760
 ctcttcagac catgctacaa gtaaggacct tgatcccttg gtagaggcaa agaatgaca 2820
 ttttaattgta tcaagctaaa gtacaagata aatgtactta caagctgtgc tatataaaaa 2880
 ttaacaagta aagggtgttct ggaaaagtaa gtcaggctca tttgtatttt tggatatgtc 2940
 tgattctatc tgcgctccaa aaacagggaa ttgtatcatt tgcacttctt aggttgtctg 3000
 60 aacttgaatt taccactatg gcatatgcat atgtggtaaa tcatgaaaga atgctcgcta 3060
 tagaaaaata aaaaaaggaa cattttttta gtgtacagaa atgatttcac ttttttctta 3120
 ttgcttaaga aaagttaaa actcttcttt gttttccttt ttgaatggta ccttcaaatg 3180
 atccattatt aagagggtta ctttttagtct ccctttgatc aggggaaatg tgcaaaactt 3240

	tatatcaaga	ttcaagtcaa	catttttga	aatgtgaag	ttctttgttt	3300
	gtctccctac	tccattaccg	acagactggc	ttcttttttg	gtttgatttc	3360
	agtctagtga	gggcaagcat	tttcttatca	gcaattctct	ttcaacaatc	3420
	aatgtaggtg	ctgttttttg	gggatatttt	ggtagtagta	cactatgttc	3480
5	ctgaaaatgc	ctcaattgat	gccttttcat	ttttcaagat	ggttttgtat	3540
	aatagtttca	gctgtgggaa	atcagtttaag	tatcactctt	cacacatctc	3600
	gcagcaaaaa	gctgatggaa	caggcaactc	tctatcattc	ttactggatg	3660
	tggctctgatt	tttgtggtaa	aacttgatga	gtatacttgg	ttgagaaatg	3720
	atttgatttc	tacattcctc	ttacaagagc	agacgccgag	gctgacaaac	3780
10	tgataagtct	tcacaagatg	atggcttaat	cagtgatata	aggaatctgg	3840
	gtcgtctaga	agaggtcagc	gagcgaagaa	taaagttctg	caagaggata	3900
	aattgagagg	ttagcggcag	aagcttatag	cattttttagg	agccccacaa	3960
	agaggaatct	gtttacattg	atgactcatc	cattgtgaag	cctgcttgtt	4020
	atctggattt	gaaatattgt	gtcaaggatt	taactgggaa	tctcataagt	4080
15	gtatgtggaa	cttggctcaa	aggccaagga	gttgtcatcg	atgggtttca	4140
	gtcaccacca	cctactgatt	ctgtgtcgcc	tgaaggatac	atgccaaagg	4200
	tctaaattcc	aggtatatct	ctgtgtgcac	aattattgtg	ttacttttagt	4260
	atgaaatttg	aaatattagt	aataattggtg	ctaagctgga	agttttacact	4320
	acaaaaaat	gagatgggat	agttggtaag	ttagagataa	agctagtaac	4380
20	ctaagatggg	gctatgctcc	agcttgtttt	atgtaaagcc	atgctgtgta	4440
	ttatgatgga	taatcagtct	accctacctt	aagtttcctt	cttctccggt	4500
	cttctcctt	gctgtttcca	aaccaatcta	ctgctaattc	tccattgcct	4560
	accttctaca	atctagcata	tagcacttgt	gctgaagatt	ttagtactac	4620
	agaaacgaaa	acctgggtag	catatgaaat	tcacatatgg	attctaaagg	4680
25	ttttataccc	ctaaaatctc	tattaaaggc	actattactg	ctgcagtagc	4740
	tctatggtta	acactaccac	cactatgcaa	taaatattta	tatgtcgctt	4800
	ttgtccacta	tgatgcatgg	atacgcgata	ctggtaacgg	gatacgagat	4860
	acatcaatac	gtcgatacgg	cgagtatata	taaaaatatg	tttaaaaaata	4920
	aactacaaaa	gggtaaagca	ataagattat	cttaaaaatat	attctatgaa	4980
30	ttacaaaaaa	tatagcaaat	atagccttgt	attacaaata	tagcaagtat	5040
	taaatatatc	agaagtatag	cattgggtct	ttaatgtgct	attgaacaag	5100
	tgctaggtat	caaagaagta	tctgataagt	atcgtgtgag	tatcggtatt	5160
	tttttttcca	ttgtttaatg	accgatactc	ctccgatacc	gtatcgacag	5220
	gtatcgccgt	atcggatacc	ggtaacggca	gaaattgaag	tatcggtgca	5280
35	tccagaaaac	tggctctcat	cctacttgat	tttgactacc	cttatgaagc	5340
	ttgtcttctc	tgtttccaaa	aaaaaaaaaa	aaacaatgat	tttcaaaggg	5400
	cactgaaaga	tctatgcctt	tatacagcat	aatagtatga	acccatcatg	5460
	ccattcatta	atatgatata	atcaatgttt	cagatatggg	accatggaag	5520
	ggctgtgaaa	cgttttcatg	aagccgggat	gaaggttctt	ggatgatgcc	5580
40	cagggtgtgct	caatttcaga	acaaaaatgg	cgtctggaat	atttttgggtg	5640
	ctgggatgat	cgagcagttg	ttgcagatga	tccacatttc	cagggtgaggt	5700
	cttgcaaaact	atttaattag	aattatttga	aaatgtcatt	catgaatgtt	5760
	actaagtatc	acaacattat	gatttttctt	atgaactgct	ggtttcatth	5820
	tttcctccag	tggagcttga	taaagttatg	cgaaaataag	cttaatttta	5880
45	atgcttggag	atagtctcat	tgagtgaact	acaaactttg	atagactcca	5940
	ttgcatacat	ttagtattgg	aagtaaacac	agattagcat	ccgaaataac	6000
	acatagtctc	aaaatggaaa	aaagacatac	tcaaacttca	gcagccatga	6060
	atcataatat	tgttctacat	gcctattcca	tttcatttta	cttaacaatg	6120
	tgtactagac	ttactttcta	atttttgggt	aaaactatat	gtcctttagt	6180
50	acaagagcag	tggagataac	ttccatgcag	ccccaaacat	tgatcactcg	6240
	tgaggagtga	tcttaaagaa	tggctttgtt	ggatgaggta	tgotatcttg	6300
	gctaaaactc	cttttggtgc	tcatgcttaa	ctatatatgt	gtattcccca	6360
	tgtatgttat	agcatcccca	tgcaacagat	taaatagatt	tgatgtacat	6420
	tggatggcgt	gttttagctga	gtgctagtgt	tctttcctca	tatcacattc	6480
55	tatgttttca	tatatgttag	tttaccttgc	tcttaccaca	aaatggaatg	6540
	acatcctcat	caatatcaca	agattccttt	tcatgctacg	ataataggta	6600
	gctgctaagc	tcctataata	acaaactgga	ttggacaaat	gctacacctg	6660
	aattgtcatg	ggttccacca	tatatatatg	tatatatat	tttgccatgt	6720
	tgatcgattc	tgtatctatg	taattgtgta	gtgcttggac	atatgcaaaa	6780
60	gatattgcag	gagctttacc	ctcattaata	ctgaaactat	ttcttggagc	6840
	gggagtattg	ttgggaacag	tagcttacta	gtcctaattg	ggttgctgca	6900
	agctgaacag	gataaattaa	atcatagaac	atagcattag	agcagcattc	6960
	tagcatcaga	tgacttcatt	gtatgtttgc	ttctgcttga	cttgcaaaaa	7020

	tacgatggat	ggcgacttga	ttttgttcgc	ggatttttggg	gtggatatgt	ccacgattac	7080
	ttggaagcaa	gcgaaccata	ttttgcagta	ggagagtact	gggattctct	cagttacacc	7140
	tatgggtgaaa	tggattataa	tcaagatgcc	cacaggcaga	gaatagttga	ttggataaat	7200
	gctacaaatg	gaactgctgg	tgcatttgat	gttaccacga	aaggaatact	tcactctgtg	7260
5	agactccatt	ttcttaccac	gtttttataat	agtgccttatg	ggtgatttag	tatttttgaa	7320
	actctagaat	ttatatgcct	ttccatgggt	tgagcttctc	agttcggata	tgcatcatga	7380
	agaatttcct	gcccttttagc	atttttcttt	tgggtggagg	tatactttgt	ccttcaatta	7440
	tttgataatt	aaagttttat	tgctgctatg	aagacgcagc	atgctgaacc	acattaatta	7500
	aacttgcatc	tttggcagta	ttcctgttcc	atggcaatat	ttaaagtctc	attaatctgt	7560
10	tgtaaagtgg	ttttctgatg	atagtttaact	catccatgat	tgtaatagcc	aattctgcca	7620
	gtttgagaag	gatgctattg	ctctggatta	gccatgcata	ttttgtctta	cacatctttt	7680
	ctggtataca	ggcactggaa	agatctgagt	actggcgtct	gtctgatgaa	aaaggaaaac	7740
	cccctggagt	gttaggttgg	tggccttcgc	gtgctgtcac	atztatagaa	aatcatgaca	7800
	ctggttctac	tcaggtagat	tacaaaatcc	ctacatacac	aaataagtat	tactccgtag	7860
15	cattatcttc	agaaggttac	tatatggcag	catgattctg	gcaaatctga	catcataatt	7920
	ttgttatatt	tgcatagaaa	catcacaaat	gtgcaacgac	ctgcatgcat	ttcttaataa	7980
	tagtcagata	gttgtccctt	tttttgaaacg	ctgcttaatt	tcttgttttc	aaacaactaa	8040
	attttgctca	gcccattcctt	gtagaactct	tttgaaactg	gaatgctttt	atggatacta	8100
	atcaaattat	aactggagct	gttaagctgt	gttcgaacaa	aatcatggct	aactttcttt	8160
20	gatatctttg	tagggtcatt	ggagattccc	ctttggtagt	gagttgcaag	gctatgtcta	8220
	catcttaact	caccagggca	ctcctgcaat	cttctatgat	catataattt	cgcatttaca	8280
	gccagagatt	gctaaattaa	tttctattag	aaatcgccaa	aagatccatt	gccgtagcaa	8340
	ggtaaaaactt	cctcttgtat	cgtatggtac	cccctcttac	cctccttgcc	cttccccact	8400
	ctattcatgt	tctctaattg	ttgaatttat	ataacattta	cagatcaaga	tactgaaagc	8460
25	agagggaat	ttatatgcgg	cagagattga	tgagagggtg	acaatgaaga	ttggcgagg	8520
	acattttgag	ccaagcggcc	ccacaaactg	ggtagttgct	gccgaggggac	aggattacaa	8580
	ggtctgggaa	gtgtcatcgt	ag				8602
	<210> 55						
30	<211> 4486						
	<212> DNA						
	<213> Arabidopsis thaliana						
	<400> 55						
35	atgtccactg	ttcccattga	gtctcttctc	caccattctt	atcttcgtca	caactcaaaa	60
	gtcaatcgcg	gaaatagaag	cttcatacct	atctcggtga	atctccgttc	tcatttcact	120
	tctaacaaac	tactacactc	aattggaaaa	agtgtcgggtg	ttagctcgat	gaacaaaagt	180
	cccgtcgcca	ttcgtgctac	ctcatcggat	actgccgtcg	tggaaaccgc	tcaatcggat	240
	gatgtcatct	tcaaggaaat	tttccctggt	cagcgaatcg	aaaagggtttt	ctcaatcgcc	300
40	gtacacagtc	acatttagcg	gaaattgatt	gaatatagtt	cctatgttgt	tgtagtttaa	360
	atctaataaa	cacattgatt	ttgtgtatgt	gttttgctga	atttaattgg	tttgatttcg	420
	atttttggga	atgataggca	gaaggaaaga	tttatgttcg	attaaaggaa	gtgaaggaga	480
	agaattggga	gctgagtgtt	ggatgtagta	tacctggaaa	atggattttg	cattggggag	540
	tttcatatgt	gggtgatact	ggcaggctct	ttcctcattt	acctcttttt	tttggttgct	600
45	tcattttacat	tattgagaaa	attttattcc	cttgatggct	tctacatttt	cgattcgata	660
	gtgaatggga	tcaacctccg	gaagatfatga	gacctcctgg	ctcaattgcc	attaagggtat	720
	tactctgggtg	tcttttttcc	cttttggttt	atatctactt	tggatagttg	taatatgttg	780
	ttatgtgggtg	tatcaggact	atgccataga	gacacctttg	aagaagttaa	ctgaaggaga	840
	ttcttttttt	gaagtgtgta	ttaatctaaa	tctggagagt	tctgtagcag	ctctcaattt	900
50	tgttttgaag	gttctgctgc	ttttaacctt	tgtaacaaag	ttgattgttg	gtttggcttt	960
	agttgtgcat	ttatttggga	ctgttttgtt	tccaggacga	agaaactggg	gcgtgggtatc	1020
	agcaciaaagg	gagagacttt	aaggttcctc	ttgtagacga	tggttcctgat	aatggaaatt	1080
	tgatcggagc	taagaaagga	tttggttaaat	ggcctttttt	ttatatatat	atatttgtca	1140
	atagcttcta	ctattgtccg	gactacagat	tttgcatctc	tatagttagg	cattctttca	1200
55	cctcctcatt	cttttttaac	gatgtataat	gacggaagtt	agtattgcaa	cagactttct	1260
	ttgaatcttt	aaactgatta	aggcttgctt	tctcttggtg	ctgtagggtg	ccttgggcag	1320
	ctttcaaaca	tccctctcaa	acaagataag	tccagtgcag	agactgattc	tattgaagaa	1380
	agaaaagggtc	ttcaagagtt	ttacgaggag	atgccaataa	gtaaacgtgt	tgccgatgat	1440
	aactcagtc	gtgtcactgc	aaggaaatgc	cctgaaacat	ctaagaacat	tgtatcaatt	1500
60	gaaaccgatt	taccgggtga	cgttactgtc	cactggggag	tttgcaaaaa	cggcactaag	1560
	aaatgggaaa	ttccatctga	accttaccct	gaagagacat	ctttgtttta	gaacaaggca	1620
	ttacgcactc	gtttacaggt	acatgtagaa	gtaggcaaga	gtcacatgtt	cagatttgaa	1680
	aatttggcaa	gaatcacttg	tggtgtgtct	ctgtaatttt	agtctgtgag	attcaaaatc	1740

	agttttttggt	tagcacctgg	cgtattcgca	attgtccaag	caatthtgtag	gttgtgcagg	1800
	cctggthtttt	aattaagcat	tagaatgggt	agcaacccaa	tgcattcggg	tcttctgttt	1860
	tttgcgccacc	ctgatagaca	tttccttcat	actthttacca	catgatgccg	tcactatgtc	1920
	tttgctgctg	cttccatata	tctthttgggc	acctagagat	cattthtcatg	ttggcaccta	1980
5	aatathtttgt	ctagcttacc	tattagaata	thttctgtgt	ctatcagtga	cattaccta	2040
	gtgtthtatgc	atatgaacat	tctgcagcga	aaagacgatg	gaaatggatc	gtttggatta	2100
	ttctctctgg	atggaaagct	tgaagggtta	tgctthgttc	ttaagthaaa	tgaaaatact	2160
	tggctaaatt	atagggggga	agacttctat	gtccctthtc	ttacttcaag	tagctcgccc	2220
	gttgaaactg	aagctgcccc	agtgaataaa	cccaaaccga	aaacagataa	agaagtgtct	2280
10	gctagtggat	ttactaaaga	aatcatcacg	gagataagga	acttggcaat	tgacattthc	2340
	tctcataaga	atcagaagac	aaacgtcaaa	gaagtgcagg	aaaacattct	acaagaaatt	2400
	gagaaactgg	ctgcggaggc	atatagcata	thtagaagca	caactccagc	thttthccgag	2460
	gaaggtgttt	tagaagcaga	agctgacaag	cctgacatta	aaatctcctc	aggaaccggc	2520
	tcgggatttg	agatattatg	ccaaggthtc	aactgggagt	ccaataaatc	tgggagatgg	2580
15	tacttggaa	ttcaagaaaa	agccgatgag	thtagctthc	ttggattcac	tgthctgtgg	2640
	ttacctccac	cgacagaatc	tgtgtcacct	gaagggtaca	tgcttaaga	cctgtataac	2700
	ttgaattcca	ggtgagthtc	gcattthttat	cagactgtct	thtgtagctt	agthgcttaa	2760
	acttcgttat	ggattgttga	atcataaagt	agaattctgg	thaatgttgt	thttgttgcc	2820
	thaaatgcat	gatcagccag	cagtggaaat	gttaattctc	thaggagatt	aacaatactt	2880
20	atgcattthc	atgatatcat	gtatctgaag	tacatgtthc	cccctattth	thgttgcaaa	2940
	ctthcagata	tggaacaatt	gatgagctaa	aagatacagt	gaagaaattt	cacaaggthc	3000
	ggatcaaagt	thtaggcgat	gctgtthctga	atcacccgtg	tgcacactth	aaaaatcaaa	3060
	atgggtgtatg	gaatctattth	ggaggacgtc	thaaactggga	cgatagggca	gtagthgcag	3120
	atgacctca	ththccaggta	thactthtct	thctthtgat	thaaaataaa	cagctataga	3180
25	acgaaaacga	ggcaaattth	thgatagaga	atgtgcttht	tgctgttatc	tgctgccgta	3240
	thtatgaact	ggactagtgt	ththtgagaga	aataaggthc	atatgatctc	tattaccctt	3300
	gatgatacaa	tcaactaatt	cattgattca	gggtagagga	aacaagagca	gtggagataa	3360
	ththccatgct	gctccaaaca	tagatcactc	gcaagactth	gttaggaagg	atatcaagga	3420
	atggctatgc	tggtatgatgt	tagthccctc	gttatthgtt	thaatatctc	thctgattag	3480
30	thtggtgtct	gattthctthc	tctattthtga	ththcacctgc	agggaagaag	thgggtatga	3540
	tggtatggagg	cttgacttht	thagagggtt	ctggggagggt	tatgtgaaag	actatatgga	3600
	tgctagcaaa	ccgtacttht	cggttggtga	atattgggat	tcgttaagth	acacgtacgg	3660
	agaaatggac	tacaatcaag	acgcacatcg	tcaaagaata	gttgactgga	thaatgcaac	3720
	tagtggagct	gctgggtgcat	thgatgtcac	taccaaagga	attctthcata	cggtactaat	3780
35	aatatcgtht	cttgthtacgt	thgatacaac	attthccttht	ggtatataat	atctggtagc	3840
	ththctgtgtg	taggcgctth	aaaaatgtga	atattggaga	ctctcagacc	caaaaggga	3900
	gcctccagggt	gtagthgggt	gggtggcctc	tcgtgctgta	acattcatcg	agaatcatga	3960
	cactggctct	acacaggtht	tattthctcag	cattctaatg	acctcaagth	tctgtthctc	4020
	ataaataaac	cgtcttaaca	attthctthc	ctththctta	agggtcattg	gagattthccg	4080
40	gagggggaagg	aaatgcaagg	atatgctthc	atcctaactc	atccaggaac	accagcggtc	4140
	thctthcgacc	atatctthctc	ggattatcat	thcgagattg	ctgcacttht	ctctctcagg	4200
	aacagacaga	aactccactg	tcggagtga	gtaaagtagt	aatcaagaac	thacaatctt	4260
	tatacaaaat	thcgacgggt	ctthaaaacca	gcctccttht	tctthcctthc	ththgtgaat	4320
	tataaagggtg	aatatagaca	agagtgaag	agatgtgtat	gcggctataa	tagatgaaaa	4380
45	ggthtgcaatg	aagatcggac	cagggcatta	tgaaccacca	aacggatcgc	aaaactggth	4440
	tgtagccgth	gaaggcagag	actacaagggt	gtgggaaaca	tctthaa		4486

50

CLAIMS

1. An isolated polypeptide which comprises a plastid alpha-amylase.
2. An isolated polypeptide of claim 1 having a mature sequence derived from a plastid alpha-amylase with an N-terminal extension comprising both a plastid targeting polypeptide and a starch-binding domain.
3. An isolated polypeptide of claim 2 wherein the N-terminal extension comprises:
 - c) a plastid targeted polypeptide of an amino acid sequence selected from SEQ ID NO: 7, SEQ ID NO: 8, residues 1-70 of SEQ ID NO: 10 and residues 1-53 of SEQ ID NO: 53; or a functionally equivalent variant thereof, and
 - d) at least one repeat of the defined polypeptide motif pair:

Motif 1 y HW (G/A)y[X]₆₋₉ WXXP[X]₃₋₅PXX(T/S)

Motif 2 F(V/L)y[X]₅₋₈ W[X]₆₋₈(D/N)F
4. An isolated polypeptide of any one of claims 1 to 3 having a mature sequence derived from the *Mdamy10* amino acid sequence of SEQ ID NO:2 or a functionally equivalent variant thereof.
5. An isolated polypeptide of claim 4 wherein the mature sequence is derived from a sequence with at least 70% identity to the sequence of SEQ ID NO:2.
6. An isolated polypeptide of claim 4 wherein the mature sequence is derived from a sequence with at least 90% identity to the sequence of SEQ ID NO:2.
7. An isolated polypeptide of claim 4 wherein the mature sequence is derived from a sequence with at least 95% identity to the sequence of SEQ ID NO:2.
8. An isolated polypeptide of claim 4 comprising the mature sequence derived from the sequence of SEQ ID NO:2.

9. An isolated polynucleotide encoding a plastid targeted alpha-amylase having the amino acid sequence of SEQ ID NO:2 or a functionally equivalent variant thereof.
10. An isolated polynucleotide of claim 9 wherein the amino acid sequence has at least 70% identity to the sequence of SEQ ID NO:2.
11. An isolated polynucleotide of claim 9 wherein the amino acid sequence has at least 90% identity to the sequence of SEQ ID NO:2.
12. An isolated polynucleotide of claim 9 wherein the amino acid sequence has at least 95% identity to the sequence of SEQ ID NO:2.
13. An isolated polynucleotide of claim 9 encoding the amino acid sequence of SEQ ID NO:2.
14. An isolated polynucleotide having the sequence of SEQ ID NO:1 or a functionally equivalent variant thereof.
15. An isolated polynucleotide of claim 9 wherein the polynucleotide is at least 70% identical in sequence to the nucleic acid sequence of SEQ ID NO:1.
16. An isolated polynucleotide of claim 9 wherein the polynucleotide is at least 90% identical in sequence to the nucleic acid sequence of SEQ ID NO:1.
17. An isolated polynucleotide of claim 9 wherein the polynucleotide is at least 95% identical in sequence to the nucleic acid sequence of SEQ ID NO:1.
18. An isolated polynucleotide of claim 9 comprising all of the sequence of SEQ ID NO:1.
19. An isolated polynucleotide comprising at least 15 contiguous nucleotides of the sequence of SEQ ID NO:1.
20. An isolated polynucleotide comprising the polynucleotide sequence of SEQ ID NO:3.

21. An isolated polynucleotide comprising a sequence which encodes the amino acid sequence of SEQ ID NO:4 or a functionally equivalent variant thereof.
22. An isolated polynucleotide comprising a sequence which encodes the amino acid sequence of SEQ ID NO:4.
23. An isolated polynucleotide which encodes the amino acid sequence of SEQ ID NO:5, or a functionally equivalent variant thereof.
24. An isolated polynucleotide having the polynucleotide sequence of SEQ ID NO:6.
25. An isolated polynucleotide encoding a chimeric protein which comprises a plastid targeting peptide of an amino acid sequence selected from SEQ ID NO:7, SEQ ID NO:8, residues 1-70 of SEQ ID NO:10 and residues 1-53 of SEQ ID NO:53; or a functionally equivalent variant thereof.
26. An isolated polynucleotide encoding a chimeric polypeptide which comprises a polypeptide which comprises at least one repeat, of the defined polypeptide motif pair:

Motif 1: yHWGV[X]₇₋₁₀W(D/E)(Q/I)P(P)[X]₃₋₄P(P)[X]₈A(I/L)XTXL

Motif 2: FV(F/L/V)K[X]₂E[X]₂₋₃W[X]₄₋₆GXDF

where capital letters represent conserved amino acids; letters in parentheses represent partly conserved amino acids; y represents a hydrophobic residue; X represents any amino acid; and [X]₄₋₆ represents a run of 4 to 6 unspecified amino acids.

27. An isolated polynucleotide encoding a chimeric polypeptide which comprises at least one repeat of the defined polypeptide motif pair:

Motif 1: yHW(G/A)y[X]₆₋₉WXXP[X]₃₋₅PXX(T/S)

Motif 2: F(V/L)y[X]₅₋₈W[X]₆₋₈(D/N)F

where capital letters represent conserved amino acids; letters in parentheses represent partly conserved amino acids; y represents a hydrophobic residue; X represents any amino acid; and [X]₄₋₆ represents a run of 4 to 6 unspecified amino acids.

28. A genetic construct comprising a polynucleotide of any one of claims 7-25.

29. A genetic construct of claim 28 comprising in a 5'-3' direction

5

- (a) a promoter sequence;
- (b) an open reading frame polynucleotide encoding a polypeptide of any one of claims 1-6; and
- (c) a termination sequence.

10

30. A genetic construct of claim 29 where the encoded polypeptide the amino acid sequence comprises the sequence of SEQ ID NO:2.

31. A genetic construct of claim 29 where the open reading frame is in a sense orientation.

15

32. A genetic construct of claim 29 where the open reading frame is in an anti-sense orientation.

33. A genetic construct comprising in a 5'-3' direction

20

- (a) a promoter sequence;
- (b) a non-coding region of a gene encoding a polypeptide of any one of claims 1-6 or a functionally equivalent variant thereof; and
- (c) a termination sequence.

25

34. A genetic construct of claim 33 where the polypeptide has an amino acid sequence of SEQ ID NO:2 and the non coding region is in the antisense orientation.

35. A genetic construct comprising in a 5'-3' direction

30

- (a) a promoter sequence;
- (b) a polynucleotide comprising a polynucleotide sequence complementary to at least 15 residues of a sequence coding for a polypeptide of any one of claims 1-6, or a functionally equivalent variant; and
- (c) a termination sequence.

35

36. A host cell comprising a genetic construct of any one of claims 28-35.
37. A transgenic plant cell which comprises a genetic construct of any one of claims 28-35.
38. A method for modulating the starch content of a plant, the method comprising:
increasing or decreasing expression of a polypeptide of any one of claims 1-8;
wherein said increasing or decreasing is achieved by genetic modification to alter the
expression of a gene encoding a plastid targeted alpha amylase.
39. A method as claimed in claim 38 wherein the polypeptide comprises a polypeptide
with the sequence of SEQ ID NO:2.
40. A genetic construct comprising a polynucleotide sequence comprising the plastid
targeting signal of SEQ ID NO:7.
41. A method for altering starch metabolism in a plant, the method comprising of
(a) introducing into the plant, a genetic construct of claims 28-35; and
(b) transcriptionally expressing the polynucleotide in the plant.
42. A method for altering starch metabolism in a plant, the method comprising of
(a) introducing into the plant, a DNA genetic construct of claims 28-35; and
(b) expressing the polypeptide in the plant.
43. A genetic construct comprising in a 5'-3' direction
(a) a promoter sequence;
(b) a polynucleotide sequence encoding a plastid targeting signal polypeptide of
claim 25 or a functionally equivalent variant thereof, translationally fused to an
additional polypeptide-encoding polynucleotide; and
(c) a termination sequence.

- 5
44. A genetic construct of claim 43 in which the polynucleotide sequence, encoding a plastid targeting signal polypeptide, comprises that of SEQ ID NO:6 or a functionally equivalent variant thereof.
45. A genetic construct of claim 44 in which the plastid targeting signal polypeptide of the invention, comprises that of SEQ ID NO: 7 or a functionally equivalent variant thereof.
- 10
46. A host cell comprising a genetic construct of claims 43-45.
47. A transgenic plant cell which comprises a genetic construct of claims 43-45.
- 15
48. A genetic construct comprising a polynucleotide of claim 26 or 27 encoding a starch binding domain polypeptide.
49. A genetic construct comprising in a 5'-3' direction
- 20
- (a) a promoter sequence;
- (b) a polynucleotide sequence encoding a starch binding domain polypeptide of claim 26 or 27, or a functionally equivalent variant thereof, translationally fused to an additional polypeptide-encoding polynucleotide; and
- a termination sequence.
- 25
50. A genetic construct of claim 49 in which the polynucleotide sequence, encoding a starch binding domain polypeptide, comprises a sequence of SEQ ID NO: 5, and functionally equivalent variants thereof.
- 30
51. A host cell comprising a genetic construct of claims 48-50.
52. A transgenic plant cell which comprises a genetic construct of claims 48-51, as well as a transgenic plant comprising such cells.
- 35
53. A plant genetically modified so that it does not express a peptide selected from SEQ ID NO:4, and SEQ NO:5; or a functionally equivalent variant thereof.

54. A method as claimed in claim 41 or 42 wherein the genetic construct comprises a
5 polynucleotide selected from SEQ ID NO:1, SEQ ID NO:15 and SEQ ID NO:52, or a
functionally equivalent variant thereof.

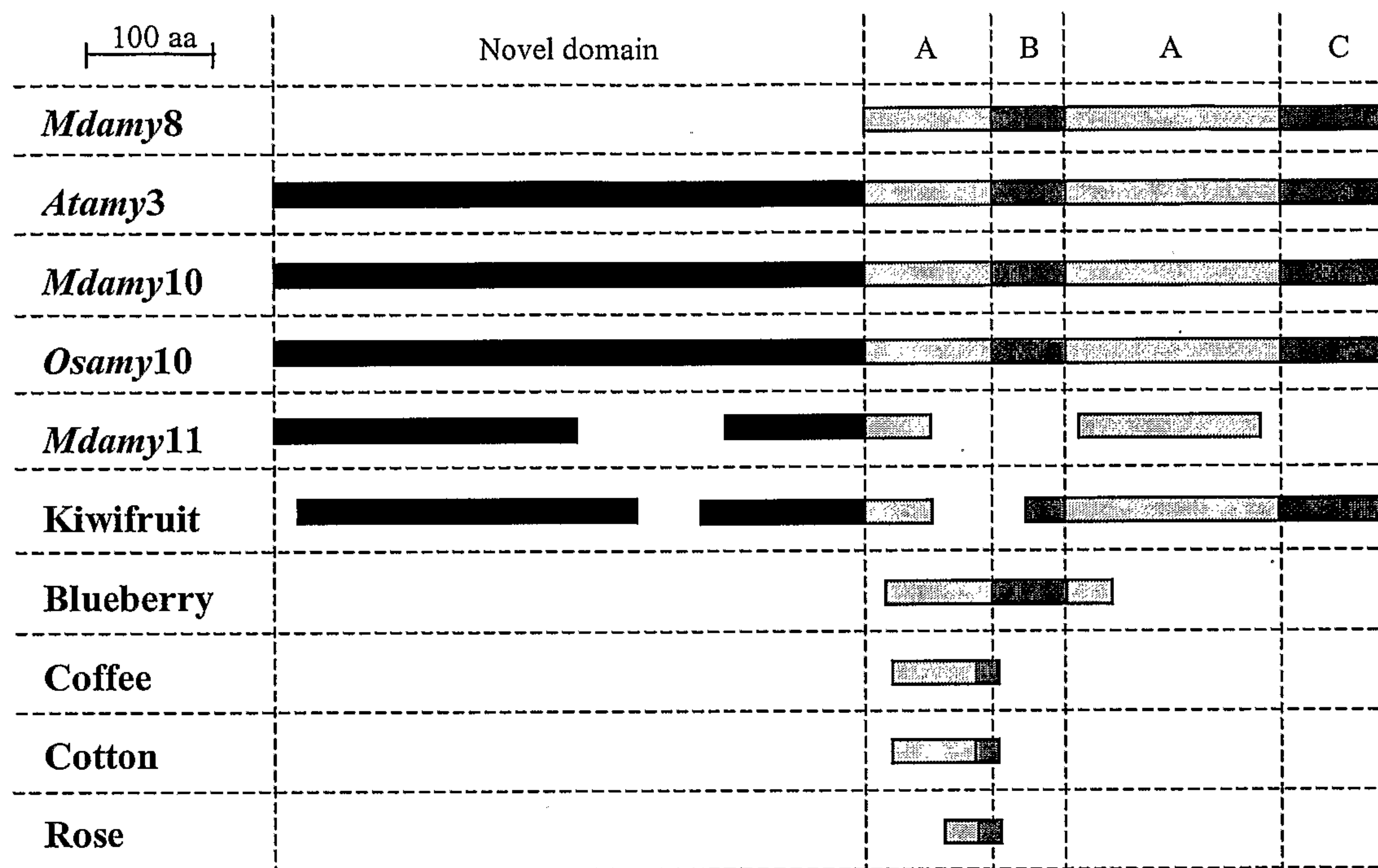
FIGURE 1

FIGURE 2

SEQ ID NO:1

AMY10:

Length: 3324 March 19, 2002 14:44 Type: N Check: 2469 ..

1 ATTCCGTCCA TCAGTTTCGT TTCTTCAGTC GCCTGAGCGG CCGTTGCTCG
51 TTCCATGTCG ACGGTTAGGA TAGAGCCCCT CCTCCACCAC TACCGTCGAC
101 AGAAACCCAG CCACCGCCTT CCGCCGTCGA AGCACCCATT AAAGCTCAGC
151 TCTTCTTTCA CTGCTTTTCC AAAGAAGCTA GTAGTCTCCA ATGGCCGCAG
201 CTTCTGCAAC TTCCAGCCTC CCACTCTCAG TGTGAGAGCT GCCTCCACCG
251 ATACAGCCAC CGTCGAGGCC ACCGAATTCG CCGACGCCTT CTACAAGGAG
301 ACCTTCCCTC TCAAGCGAAC TGAAGTGGTG GAGGGAAAGA TGATCGTGAA
351 ATTAGATAAT GGGAAGGATG CAAAGAATTG GGTGCTGACT GTGGGTTGTA
401 ATCTTCCTGG AAAATGGGTT CTTCACTGGG GAGTTAATTA TGTGGATGAC
451 GTCGGCAGTG AATGGGATCA GCCTCCTAGT GAAATGAGAC CAGCTGGTTC
501 AGTTTCCATC AAGGACTATG CAATAGAGAC ACCCTTGAAG GAATCGTTGT
551 CGCCGGTGGG AGGCGATACA TCTCACGAAG TGAAGATTGA TGTTACACCC
601 AACAGCGCAA TCGCAGCGAT AAATTTTGT TCAAGGATG AAGAAACCGG
651 TGCCTGGTAT CAGCATAGAG GGAGAGACTT TAAAGTGCCT TTCGTGGGCT
701 ACCTGCAAGA CGATGACAAT GTAGTTGGAG CAACAAGGGC CTTGGGCGCG
751 TGGTCAGGAA CTTTGGGAAA ACTATCCAAT GTGTTTGTCA AAGCAGAAAC
801 ATCAAATTCC AAAGATCAAG AAAGCAGCAG TGAATCCAGA GACCCTCAAC
851 AGAAACTAT GCGTCTAGAA GGATTCTATG AAGAACTGCC AATTGCAAAA
901 GAAATTGCTG TTAACCATTC AGCAACTGTT TCCGTTAGGA AGTGCCCTGA
951 GACTACTAAG AATCTTCTAT ACTTGGAAC AGATTACCT GATCATGCTG
1001 TTGTTCACTG GGGAGTTTGC AGAGATGATG CTAAAAGATG GGAAATTCCA
1051 GCTGCCCCGC ATCCACCAGA AACAGTAGTT TTCAAGGACA AGGCTTTGCG
1101 GACTCGATTA CAGCAAAGGG AGGATGGAAA TGGATGTTCT GGACTGTTTA
1151 CCTTGGAAGA AGGACTCGCA GGATTTCTTT TTGTTTTCAA ACTAAATGAA
1201 ACTATGTGGT TGAATTGTGT GGGCAATGAC TTCTACATCC CCCTTTTAAG

1251 CTCAAATAAC TCAATTGCTG TGCAAAATGA GGTTCACTCT GAAGATGCTC
 1301 AGGTACCTGA CAGAAGTAGA GAAACTAATT TTACTGCATA TACCGATGGT
 1351 ATAATCAATG AAATAAGGAA CTTGGTGAGT GATATTCCT CTGAGAAGAG
 1401 TCAAAGGAAA AGATCCAAAG AAGCACAAGA AACCATTCTT CAAGAAATAG
 1451 AAAAATTGGC TGCTGAAGCG TATAGTATCT TCAGAACTAC TGTTCCAAC
 1501 CTACCGGAGG AAATCATTGC AGAACTGAA AAGGTGAAAG TCGCCCCTGC
 1551 CAAAATATGC TCGGGGACAG GAACAGGTTT TGAAATATTG TGCCAAGGTT
 1601 TTAAGTGGGA ATCTAGTAAA TCTGGAAGAT GGTACGAGGA ACTCAAAAGT
 1651 AAAGCTGCAG AATTATCCTC ACTAGGTTTC ACTGTGATTT GGTTCCCACC
 1701 TCCTACAGAT TCTGTGTCGC CTCAAGGGTA CATGCCGAGG GATTTATACA
 1751 ATATGAACTC CAGATATGGA AATATGGACG AACTGAAGGA GACTGTGAAG
 1801 ACATTCCATG ATGCCGGTTT AAAAGTTCTT GGAGATGCTG TTCTGAATCA
 1851 CCGTTGTGCA GAATATCAGA ATCAAAATGG TGTTTGGAAT ATATTTGGTG
 1901 GTCGTTTAAA TTGGGATGAA CGTGCAGTTG TTGCAGATGA TCCACATTTT
 1951 CAGGGTAGGG GCAACAAAAG TAGTGGAGAT AGTTTTCATG CTGCCCCAAA
 2001 CATTGATCAT TCACAAGATT TTGTGAGGAA GGATATCAGA GAATGGTTAT
 2051 GCTGGCTAAG GGACGATATT GGGTATGACG GATGGAGGCT TGATTTCTGT
 2101 AGAGGATTTT GGGGTGGCTA CGTCAAGGAC TACATGGATG CCAGTGAGCC
 2151 CTACTTTGCC GTAGGCGAGT ATTGGGATTC CCTAAGTTAT ACATACGGGG
 2201 AAATGGATCA CAATCAAGAT GCACACAGGC AGAGAATTGT TGATTGGATC
 2251 AACGCTACTA ATGGAAGTTG TGGTGCATTC GATGTCACAA CAAAAGGGAT
 2301 TCTCCATGCA GCACTGGAAA GATGCGAGTA TTGGCGATTG TCAGATGAGA
 2351 AGGGTAAGCC TCCAGGGGTT CTTGGATGGT GGCCATCTCG TGCTGTCACT
 2401 TTCATAGAGA ATCATGATAC TGGTTCTACT CAGGGTCATT GGAGGTTTCC
 2451 AAATAAAAAA GAAATGCAAG GATATGCATA CATTTTGACT CATCCGGGAA
 2501 CACCTACAGT TTTCTATGAC CACATTTTCT CTCCTACCA ATCTGAAATT
 2551 GCTGCTCTGA TCTCTCTCAG AAACCGGAAC AAGCTCAACT GTCGGAGTAG
 2601 AGTTAAAATC ACCAAGGCAG AAAGAGACGT CTACGCGGCC ATCATTGATG
 2651 AAAAGGTCGC CATCAAAATC GGACCAGGTC ATTACGAACC TGCTAGTGGA
 2701 CCTCAAAATT GGAATAAAAG CCTTGAGGGA AGAGACTACA AGGTCTGGGA

2751 AGCGTCATAA ATTTTTTCTG CACCGGTATG TATGTAACCTT CGAGTGTATA
2801 AGGTTTGAGT ACCGGATGAT TCAATAAGGA AGTTCCCATA TTTACATTGT
2851 AAGCAAATCA AAGGATGCCG AGATAGAGTA ATATGTACTT CTCTCTAGAG
2901 GTCTGGCTGA CACCCTTGAA TTGAAGTTGT TTCGGCAGAC ATTTTCCGAA
2951 TTTAAGGGAC TGTAGACGCT CAACTGAGGA CGAGACGGGC CAGCCTGCTT
3001 CAGCAGCACA GAATCGCGGC CAGTCGGCTG AGTGTGGAAG TGGTGAGCAT
3051 GGTACTGGGT TCCAAGGGCA GTGCTCCATC TTACACACAC TGTGACTCT
3101 CCAACTTATA CTGATGCTTT TAACCAATAA TGTATAAATT GGTGCATGCC
3151 AGTGACTCGT GGTCCATGAA CGTTGAGTGA GTTTCCTATC TGAACCTTTA
3201 TTCTGTCTCA CATTGCTCCA GTCCATCAAC TTTATGCATG GACTCGTTAA
3251 TTTTTTAGTA TTGTTCTTGA AACAGGATGA TCGATATATC GTTAATATGT
3301 GATCCATATA TTATATTAAA AAAA

FIGURE 3

SEQ ID NO:2

amy10h.pep Length: 902 March 19, 2002 15:04 Type: P Check: 1075 ..

1 MSTVRIEPLL HHYRRQKPSH RLPPSKHPLK LSSSF~~TAF~~PK KLVVSNGRSF
51 CNFQPP~~TL~~SV RAASTD~~TAT~~V EATEFADAFY KET~~FPL~~KRTE VVEGKMIVKL
101 DNGKDAKNWV LTVGCNLP~~PG~~K WVLHWGVNYV DDVGSEWDQ~~P~~ PSEMRPAGSV
151 SIKDYAIE~~TP~~ LKESLSPVGG DTSHEVKIDV TPNSAIAAIN FVLKDEETGA
201 WYQHRGRDFK VPFVGYLQDD DNVVGATRAL GAWSGTLGKL SNV~~FV~~KAETS
251 NSKDQESSSE SRDPQOKTMR LEGFYEE~~LPI~~ AKEIAVNHSA TVSVRKCPET
301 TKNLLYLETD LPDHAVVHWG VCRDDAKRWE IPAAPHPPET VVFKDKALRT
351 RLQQREDGNG CSGLE~~FT~~LEEG LAGFLFVFKL NETMWLNCVG NDFYIPL~~L~~SS
401 NNSIAVQNEV QSEDAQVPDR SRET~~NFT~~AYT DGIINEIRNL VSDISSEK~~SQ~~
451 RKRSKEAQET ILQEI~~EK~~LAA EAYSIFRTTV PTLPEEIIAE TEKVKVAPAK
501 ICSGTGTGFE ILCQGFNWES SKSGRWYEEL KSKAAELSSL GFTVIWFPPP
551 TDSVSPQGYM PRDLYNMNSR YGNMDELKET VKTFHDAGLK VLGDAVLNHR
601 CAEYQNQNGV WNIFGGRLNW DERA~~VV~~ADDP HFQGRGNKSS GDSFHAAPNI
651 DHSQDFVRKD IREWLCWLRD DIGYDGWRLD FVRGFWGGYV KDYMDASEPY
701 FAVGEYWDSL SYTYGEMDHN QDAHRQ~~RIV~~D WINATNGTCG AFDVTTKGIL
751 HAALERCEYW RLSDEK~~GK~~PP GVLGWWPSRA VTFIENHDTG STQGHWRFPN
801 KKEMQGYAYI LTHPGTPTVF YDHIFSHYQS EIAALISLRN RNKLNCRSRV
851 KITKAERDVY AAIIDEKVAI KIGPGHYEPA SGPQNWNKSL EGRDYKVWEA
901 S*

FIGURE 4

ATAMY3 :

1 ATGTCCACTG TTCCCATTGA GTCTCTTCTC CACCATTCTT ATCTTCGTCA
 51 CAACTCAAAA GTCAATCGCG GAAATAGAAG CTTCATACCT ATCTCGTTGA
 101 ATCTCCGTTT TCATTTCACT TCTAACAAAC TACTACACTC AATTGGAAAA
 151 AGTGTCTGGT TTAGCTCGAT GAACAAAAGT CCCGTCGCCA TTCGTGCTAC
 201 CTCATCGGAT ACTGCCGTCG TGGAAACCGC TCAATCGGAT GATGTCATCT
 251 TCAAGGAAAT TTTCCCTGTT CAGCGAATCG AAAAGGCAGA AGGAAAGATT
 301 TATGTTTCGAT TAAAGGAAGT GAAGGAGAAG AATTGGGAGC TGAGTGTGTTG
 351 ATGTAGTATA CCTGGAAAAT GGATTTTGCA TTGGGGAGTT TCATATGTGG
 401 GTGATACTGG CAGTGAATGG GATCAACCTC CGGAAGATAT GAGACCTCCT
 451 GGCTCAATTG CCATTAAGGA CTATGCCATA GAGACACCTT TGAAGAAGTT
 501 ATCTGAAGGA GATTCTTTTT TTGAAGTTGC TATTAATCTA AATCTGGAGA
 551 GTTCTGTAGC AGCTCTCAAT TTTGTTTTGA AGGACGAAGA AACTGGGGCG
 601 TGGTATCAGC ACAAAGGGAG AGACTTTAAG GTTCCTCTTG TAGACGATGT
 651 TCCTGATAAT GGAAATTGA TCGGAGCTAA GAAAGGATTT GGTGCCCTTG
 701 GGCAGCTTTC AAACATCCCT CTCAAACAAG ATAAGTCCAG TGCAGAGACT
 751 GATTCTATTG AAGAAAGAAA AGGTCTTCAA GAGTTTTACG AGGAGATGCC
 801 AATAAGTAAA CGTGTGCGG ATGATAACTC AGTCAGTGTC ACTGCAAGGA
 851 AATGCCCTGA AACATCTAAG AACATTGTAT CAATTGAAAC CGATTTACCG
 901 GGTGACGTTA CTGTCCACTG GGGAGTTTGC AAAAACGGCA CTAAGAAATG
 951 GGAAATTCCA TCTGAACCTT ACCCTGAAGA GACATCTTTG TTTAAGAACA
 1001 AGGCATTACG CACTCGTTTA CAGCGAAAAG ACGATGGAAA TGGATCGTTT
 1051 GGATTATTCT CTCTGGATGG AAAGCTTGAA GGGTTATGCT TTGTTCTTAA
 1101 GTTAAATGAA AATACTTGGC TAAATTATAG GGGGAAGAC TTCTATGTCC
 1151 CTTTCCTTAC TTCAAGTAGC TCGCCCGTTG AACTGAAGC TGCCCAAGTG
 1201 AGTAAACCCA AACGAAAAAC AGATAAGAA GTGTCTGCTA GTGGATTAC
 1251 TAAAGAAATC ATCACGGAGA TAAGGAACTT GGCAATTGAC ATTCCTCTC
 1301 ATAAGAATCA GAAGACAAAC GTCAAAGAAG TGCAGGAAAA CATTCTACAA

1351 GAAATTGAGA AACTGGCTGC GGAGGCATAT AGCATATTTA GAAGCACAAC
1401 TCCAGCTTTT TCCGAGGAAG GTGTTTTAGA AGCAGAAGCT GACAAGCCTG
1451 ACATTAAAAT CTCCTCAGGA ACCGGCTCGG GATTTGAGAT ATTATGCCAA
1501 GGTTTCAACT GGGAGTCCAA TAAATCTGGG AGATGGTACT TGGAAC TTCA
1551 AGAAAAAGCC GATGAGTTAG CTTCACTTGG ATTCAC TGTT CTGTGGTTAC
1601 CTCCACCGAC AGAATCTGTG TCACCTGAAG GGTACATGCC TAAAGACCTG
1651 TATAACTTGA ATTCCAGATA TGGAACAATT GATGAGCTAA AAGATACAGT
1701 GAAGAAATTT CACAAGGTCG GGATCAAAGT TTTAGGCGAT GCTGTTCTGA
1751 ATCACCGCTG TGCACACTTC AAAAATCAAA ATGGTGTATG GAATCTATTT
1801 GGAGGACGTC TAAACTGGGA CGATAGGGCA GTAGTTGCAG ATGACCCTCA
1851 TTTCCAGGGT AGAGGAAACA AGAGCAGTGG AGATAATTTT CATGCTGCTC
1901 CAAACATAGA TCACTCGCAA GACTTTGT TA GGAAGGATAT CAAGGAATGG
1951 CTATGCTGGA TGATGGAAGA AGTTGGGTAT GATGGATGGA GGCTTGACTT
2001 TGTAAGAGGG TTCTGGGGAG GTTATGTGAA AGACTATATG GATGCTAGCA
2051 AACCGTACTT TCGGTTGGT GAATATTGGG ATTCGTTAAG TTACACGTAC
2101 GGAGAAATGG ACTACAATCA AGACGCACAT CGTCAAAGAA TAGTTGACTG
2151 GATTAATGCA ACTAGTGGAG CTGCTGGTGC ATTTGATGTC ACTACCAAAG
2201 GAATTCTTCA TACGGCGCTT CAAAATGTG AATATTGGAG ACTCTCAGAC
2251 CCAAAGGGA AGCCTCCAGG TGTAGTTGGT TGGTGGCCTT CTCGTGCTGT
2301 AACATTCATC GAGAATCATG AACTGGCTC TACACAGGGT CATTGGAGAT
2351 TTCCGGAGGG GAAGGAAATG CAAGGATATG CTTACATCCT AACTCATCCA
2401 GGAACACCAG CGGTCTTCTT CGACCATATC TTCTCGGATT ATCATTCGGA
2451 GATTGCTGCA CTTCTCTCTC TCAGGAACAG ACAGAACTC CACTGTCGGA
2501 GTGAGGTGAA TATAGACAAG AGTGAGAGAG ATGTGTATGC GGCTATAATA
2551 GATGAAAAGG TTGCAATGAA GATCGGACCA GGCATTATG AACCACCAAA
2601 CGGATCGCAA AACTGGTCTG TAGCCGTTGA AGGCAGAGAC TACAAGGTGT
2651 GGGAAACATC TTAA

FIGURE 5

Atamy3:

1 MSTVPIESLL HHSYLRHNSK VNRGNRSFIP ISLNLRSHFT SNKLLHSIGK
51 SVGVSSMNKS PVAIRATSSD TAVVETAQSD DVIFKEIFPV QRIEKAEGKI
101 YVRLKEVKEK NWELSVGCSI PGKWILHWGV SYVGD TGSEW DOPPEDMRPP
151 GSIAIKDYAI ETPLKKLSEG DSFFEVAI NL NLESSVAALN FVLKDEETGA
201 WYQHKGRDFK VPLVDDVPDN GNLIGAKKGF GALGQLSNIP LKQDKSSAET
251 DSIEERKGLQ EFYEEMPISK RVADDNSVSV TARKCPETSK NIVSIETDLP
301 GDVTVHWGVC KNGTKKWEIP SEPYPEETSL FKNKALRTRL QRKDDGNGSF
351 GLFSLDGKLE GLCFVLKLE NTWLN YRGED FYVPFLTSSS SPVETEEAAQV
401 SKPKRKTDKE VSASGFTKEI ITEIRNL AID ISSHKNQKTN VKEVQENILQ
451 EIEKLAAEAY SIFRSTTPAF SEEGVLEAEA DKPDIKISSG TGSGFEILCQ
501 GFNWESNKSG RWYLELQEKA DELASLGFTV LWLPPPTESV SPEGYMPKDL
551 YNLNSRYGTI DELKDTVKKF HKVGKVLGD AVLNHRCAHF KNQNGVWNLF
601 GGRLNWDDRA VVADDPHFQG RGNKSSGDNF HAAPNIDHSQ DFVRKDIKEW
651 LCWMMEEVGY DGWRLDFVRG FWGGYVKDYM DASKPYFAVG EYWDSLSTY
701 GEMDYNQDAH RQRIVDWINA TSGAAGAFDV TTKGILHTAL QKCEYWRLSD
751 PKGKPPGVVG WWPSRAVTFI ENHDTGSTQG HWRFPPEGKEM QGYAYILTHP
801 GTPAVFFDHI FSDYHSEIAA LLSLRNRQKL HCRSEVNIDK SERDVYAAII
851 DEKVAMKIGP GHYEPPNGSQ NWSVAVEGRD YKVVWETS*

FIGURE 6

Osamy10:

```

1  ATGGCGGTGG CGAGCTGGAG CATTCGGGCG ATCCCGCGGG CGGGTCCGAC
51  AGCGAGGGGT GTGCTGCTCG GCGGCGCCTT CGTGACGGCC GCGCGGCCGC
101 CCGTGGCGTG GCGGTGCCGG GCCACGCTCC CTAGGAGAGT GAGGCTCGGT
151 GCGGTGGTGG CCCGTGCCGG TCGGCGGAG ACGCCGGTGG CCGGCTCCGG
201 GGAAGCCGGG TTGCTGTTCT CCGAGAAGTT CCCCTTGCGC CGATCCCGAA
251 CGGTGGAAGG GAAGGCGTGG GTGAGGGTCG ATGCGGAGCC GGATGGGGAG
301 GGCAAGTGCA AGGTTGTGAT CGGGTGTGAT GTAGAGGGGA AGTGGGTGCT
351 GCATTGGGGT GTCTCCTACG ATGGTGAGCA GGAAGAGAA TGGGACCAAC
401 CTCCTTCAGA CATGAGACCT CCCGGTTCAG TGCCTATTAA GACTATGCA
451 ATTGAAACAT CTTTGGACAC TCCACACAAT TCAGAAGGCA AGACGATTCA
501 TGAAGTGCAA ATCAAAATTG ATAAGGGCAC ATCAATTGCT GCTATCAATT
551 TTGTTCTAAA GGAAGAGGAA ACAGGTGCTT GGTTCAGCA CAAGGGTCAG
601 GATTTCAGAA TACCTTTAAG TGGATCCTTT GGTGGAGATC TACTAGGAAC
651 AGAACAAGAT ATTGATGTCA GGCCAGGTCA CCTATCTAAC GTGTTACAGA
701 AACCTGAGGG ACCTATTGCT GAGCCTCATA AAAGTGTACC CGATGATAAA
751 GGTTCAGAA CCAAACACAT TTCAGGTTTC TATGAGGAAT ACCCAATCTT
801 AAAAACGGTG TATGTTTACA ATTTTATAAC TGTTAATGTG AGGGAAAACA
851 ATGGAACAAC TAAACATGCT GTGGAATTCG AACTGATAT TCCTGGAGAA
901 GTTATCATTC ATTGGGGAGT TTGCAAAGAC AATACCATGA CATGGGAGAT
951 CCCCCAGAA CCACATCCAC CTGCAACGAA GATATTCCGA CAGAAAGCTC
1001 TTGAGACCAT GCTACAACAA AAAGCTGATG GAACAGGCAA CTCTCTATCA
1051 TTCTTACTGG ATGGAGAGTA TTCTGGTCTG ATTTTGTGG TAAACTTGA
1101 TGAGTATACT TGGTTGAGAA ATGTGGAGAA TGGATTGAT TTCTACATTC
1151 CTCTTACAAG AGCAGACGCC GAGGCTGACA AACAGAAAGC CGATGATAAG
1201 TCTTCACAAG ATGATGGCTT AATCAGTGAT ATAAGGAATC TGGTGGTTGG
1251 GCTGTCGTCT AGAAGAGGTC AGCGAGCGAA GAATAAAGTT CTGCAAGAGG
1301 ATATCCTACA AGAAATTGAG AGGTTAGCGG CAGAAGCTTA TAGCATTTTT
1351 AGGAGCCCCA CAATTGATAC TGTAGAGGAA TCTGTTTACA TTGATGACTC

```

1401 ATCCATTGTG AAGCCTGCTT GTTCTGGTAC TGGATCTGGA TTTGAAATAT
1451 TGTGTCAAGG ATTTAACTGG GAATCTCATA AGTCAGGAAA ATGGTATGTG
1501 GAACTTGGCT CAAAGGCCAA GGAGTTGTCA TCGATGGGTT TCACCATTGT
1551 CTGGTCACCA CCACCTACTG ATTCTGTGTC GCCTGAAGGA TACATGCCAA
1601 GGGATTTGTA TAATCTAAAT TCCAGATATG GGACCATGGA AGAGTTGAAG
1651 GAGGCTGTGA AACGTTTTCA TGAAGCCGGT ATGAAGGTTT TTGGTGATGC
1701 CGTCCTGAAT CACAGGTGTG CTCAATTTCA GAACCAAAAT GGCGTCTGGA
1751 ATATTTTTTG TGGACGCCTT AACTGGGATG ATCGAGCAGT TGTTGCAGAT
1801 GATCCACATT TCCAGGGAAG AGGAAACAAG AGCAGTGGAG ATAACTTCCA
1851 TGCAGCCCCA AACATTGATC ACTCGCAAGA GTTTGTGAGG AGTGATCTTA
1901 AAGAATGGCT TTGTTGGATG AGAAAGGAAG TTGGATACGA TGGATGGCGA
1951 CTTGATTTTG TTCGCGGATT TTGGGGTGGA TATGTCCACG ATTACTTGGA
2001 AGCAAGCGAA CCATATTTTG CAGTAGGAGA GTACTGGGAT TCTCTCAGTT
2051 ACACCTATGG TGAAATGGAT TATAATCAAG ATGCCCACAG GCAGAGAATA
2101 GTTGATTGGA TAAATGCTAC AAATGGAACT GCTGGTGCAT TTGATGTTAC
2151 CACGAAAGGA ATACTTCACT CTGCACTGGA AAGATCTGAG TACTGGCGTC
2201 TGTCTGATGA AAAAGGAAAA CCCCTGGAG TGTTAGGTTG GTGGCCTTCG
2251 CGTGCTGTCA CATTTATAGA AAATCATGAC ACTGGTTCTA CTCAGGGTCA
2301 TTGGAGATTC CCCTTTGGTA TGGAGTTGCA AGGCTATGTC TACATCTTAA
2351 CTCACCCAGG CACTCCTGCA ATCTTCTATG ATCATATATT TTCGCATTTA
2401 CAGCCAGAGA TTGCTAAATT AATTCTTATT AGAAATCGCC AAAAGATCCA
2451 TTGCCGTAGC AAGATCAAGA TACTGAAAGC AGAGGGAAAT TTATATGCGG
2501 CAGAGATTGA TGAGAGGGTA ACAATGAAGA TTGGCGCAGG ACATTTTGAG
2551 CCAAGCGGCC CCACAACTG GGTAAGTTGCT GCCGAGGGAC AGGATTACAA
2601 GGTCTGGGAA GTGTCATCGT AG

FIGURE 7

Osamy10:

1 MAVASWSIPA IPRAGPTARG VLLGGAFVTA ARPPVAWRCR ATLPRRVRLG
51 GVVARAGAAE TPVAGSGEAG LLFSEKFPLR RSRTVEGKAW VRVDAEPDGE
101 GKCKVVGCD VEGKWVLHWG VSYDGEQGRE WDQPPSDMRP PGSVPIKDYA
151 IETSLDTPHN SEGKTIHEVQ IKIDKGTSIA AINFVLKEEE TGAWFQHKGO
201 DFRIPLSGSF GGDLLGTEQD IDVRPGHLSN VLQKPEGPIA EPHKTVPDDK
251 GSRTKHISGF YEEYPILKTV YVQNFITVNV RENNGTTKHA VEFDTDIPGE
301 VIIHWGVCKD NTMTWEIPPE PHPPATKIFR QKALQTMLQQ KADGTGNSLS
351 FLLDGEYSGL IFVVKLDEYT WLRNVENGFD FYIPLTRADA EADKQKADDK
401 SSQDDGLISD IRNLVVG LSS RRGQRAKNKV LQEDILQEIE RLAAEAYSIF
451 RSPTIDTVEE SVYIDDSSIV KPACSGTGSG FEILCQGFNW ESHKSGKWYV
501 ELGSKAKELS SMGFTIVWSP PPTDSVSPEG YMPRDLYNLN SRYGTMEELK
551 EAVKRFHEAG MKVLGDAVLN HRCAQFQONQN GVWNIFGGRL NWDDRRAVVAD
601 DPHFQGRGNK SSGDNFHAAP NIDHSQEFVR SDLKEWLCWM RKEVG YDGWR
651 LDFVRGFWGG YVHDYLEASE PYFAVGEYWD SLSYTYGEMD YNQDAHRQRI
701 VDWINATNGT AGAFDVTTKG ILHSALERSE YWRLSDEK GK PPGVLGWWPS
751 RAVTFIENHD TGSTQGHWRP PFGMELQGYV YILTHPGTPA IFYDHIFSHL
801 QPEIAKLISI RNRQKIHC RS KIKILKAEGN LYAAEIDERV TMKIGAGHFE
851 PSGPTNWWVA AEGQDYKVWE VSS

FIGURE 8

Mdamy10a : LHWGVNYVDVVGSEWDQPPSEMRPAGSMSTIKDYAIEITPLKESLSPVGGDT
Mdamy11a : LHWGVSFVDVVSCEWEQPPSEMRPAGSMPIKDYAIEITPLKESLSSVGGDT
Atamy3a : LHWGVSYVGDITGSEWDQPPEDMRPPGSTAIKDYAIEITPLKKLS---EGDS
Osamy10a : LHWGVSYDGEQGREWDQPPSDMRPPGSMPIKDYAIEITSL-DTPHNSEGKT
kiwifruit : LHWGVNYINDIGSEWDQPPVEMRPPGSMPIKDYAIEITPLKKSSAVVEGDL
Mdamy10b : VHWGV--CRDDAKRWEIPA-APHPPETWVFKDKALRTRLQQREDGNGCSG
Atamy3b : VHWGV--CKNGTKKWEIPS-EPYPEETSLFKDKALRTRLQRKDDGNGSFG
Osamy10b : IHWGV--CKDNTMTWEIPP-EPHPPATKIFRQKALQTMQLQQKADGTGNSL

Mdamy10a : SHEVKIDVTPNSAIAAINFVVK-DEETGAWYQHRGRDF
Mdamy11a : SYEVKIDVKPNSAIAAINFVVK-DEETGAWYQHRGSDF
Atamy3a : FEEVAINLNLESSMAAINFVVK-DEETGAWYQHKGRDF
Osamy10a : IHEVQIKIDKGTSIAAINFVVK-EEETGAWYQHKQDF
kiwifruit : YVELKIDFSTDKDIAAINFVVK-DEETGAWYQRRGRDF
Mdamy10b : LETL-----EEGLAGFLFVFKLNETMWLNLCV--GNDF
Atamy3b : LESL-----DGKLEGICFVKLNENTWLNRYR--GEDF
Osamy10b : SELL-----DGEYSGLIFVVKLEEYTWLRNVENGDF

FIGURE 9

```

Mdamy10a : LHWGVNYVDDVGSEWDQPPSEMRFAGSVSIKDYATETPLKESLSP--VGG
Mdamy11a : LHWGVSFVDDVSCWEQPPSEMRFAGSVPIKDYATETPLKESLSS--VGG
Atamy3a   : LHWGVSYVGDVGSEWDQPPEDMRPPGSIATKDYATETPLKKLS-----EG
Osamy10a  : LHWGVSYDGEQGREWDQPPSDMRPPGSIATKDYATETPLSL-DTPHN--SEG
kiwifruit : LHWGVNYINDIGSEWDQPPVEMRPPGSIATKDYATETPLKKSSAV--VEG
Sex1a     : LHWCAILDNKE--NWVLPS-R-SPDRITQNFKNALRTPFV-KS-----GG
R1a       : LHWCAVKFGKE--TWSLPN-D-RPDGTVKYKNKALRTPFV-KS-----GS
Mdamy10b  : VHWGVCRRDAK--RWEIPA-APHPPETVVFVKDKALRTRLQQRE-----DG
Atamy3b   : VHWGVCKNGTK--KWEIPS-EPYEEETSLFKNKALRTRLQRKD-----DG
Osamy10b  : IHWGVCKDNTM--TWEIPP-EPHPPATKIFRQKALQIMLQQA-----DG
Sex1b     : LHWALSQKGGE---WLDPPSDILPPNSLPVVG-AVDTKLTITSTD-LPSP
R1b       : LHWALSQKSPGE---WMVPPSSILPPGSIILDK-AAETPFSASSSDGLTSK
SBE-like  : IHWEPSPEAAY--KWKYSK--PKVPESLRIYECHVGISGSEPKVS---TF

```

```

Mdamy10a : DTSHEVKIDVTPNSAIAAINNEVLKDEETGAWYQH--RGRDF
Mdamy11a : DTSYEVKIDVKPNSAIAAINNEVLKDEETGAWYQH--RGSDF
Atamy3a   : DSFFEVAINLNLESSVAIAAINNEVLKDEETGAWYQH--KGRDF
Osamy10a  : KTIHEVQIKIDKGTSIAIAAINNEVLKEEETGAWFQH--KGQDF
kiwifruit : DLYYELKIDFSTDKDIAIAAINNEVLKDEETGAWYQR--RGRDF
Sex1a     : NSHLKIEI---DDPAIHATEELTFDES RNKWKYKN--NGQNF
R1a       : NSILRIEI---RDTAIEATEELTYDEAHDKWIKN--NGGNF
Mdamy10b  : NGCSGIFT---LEEGLAGFLF-VFKLNETMWLNCV--GNDF
Atamy3b   : NGSFGIFS---LDGKLEGLCF-VLKLNENTWLNRY--GEDF
Osamy10b  : TGNSLSFI---LDGEYSGLIF-VVKLDEYTWLRNVENCDF
Sex1b     : VQTFEIEIE---GDSYKGMPEVLNAGER--WIKN--NDSDF
R1b       : VQSLDIVIE---DGNFVGMPFVLLSGEK--WIKN--QGSDF
SBE-like  : EEFTKKVIPHVKRAGYNALQLIGVPEHKDYFTVG-YRVTNF

```

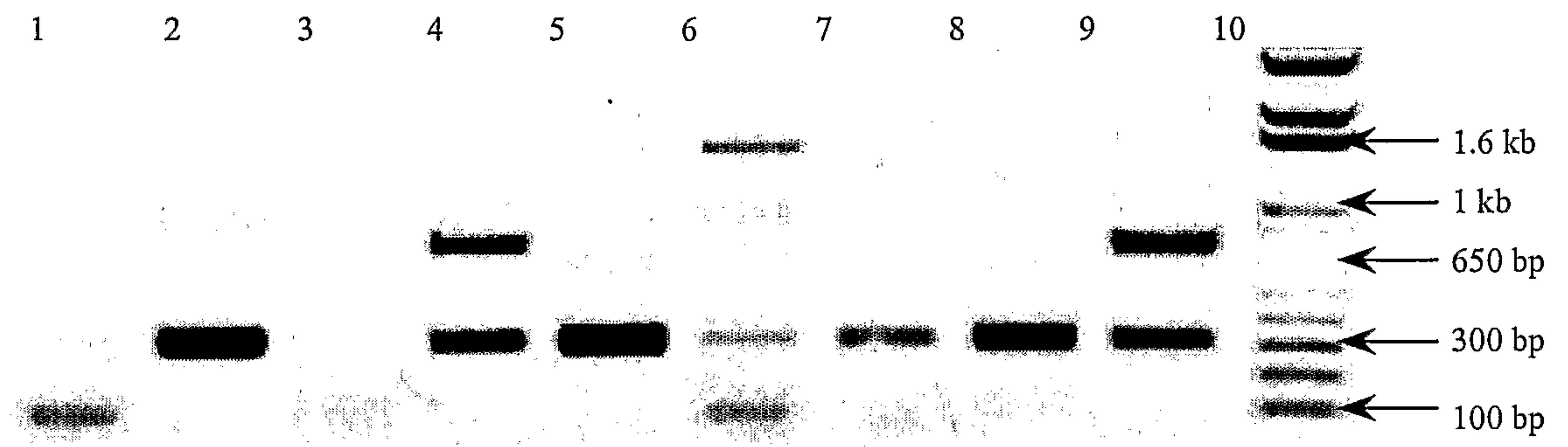
FIGURE 10

FIGURE 11

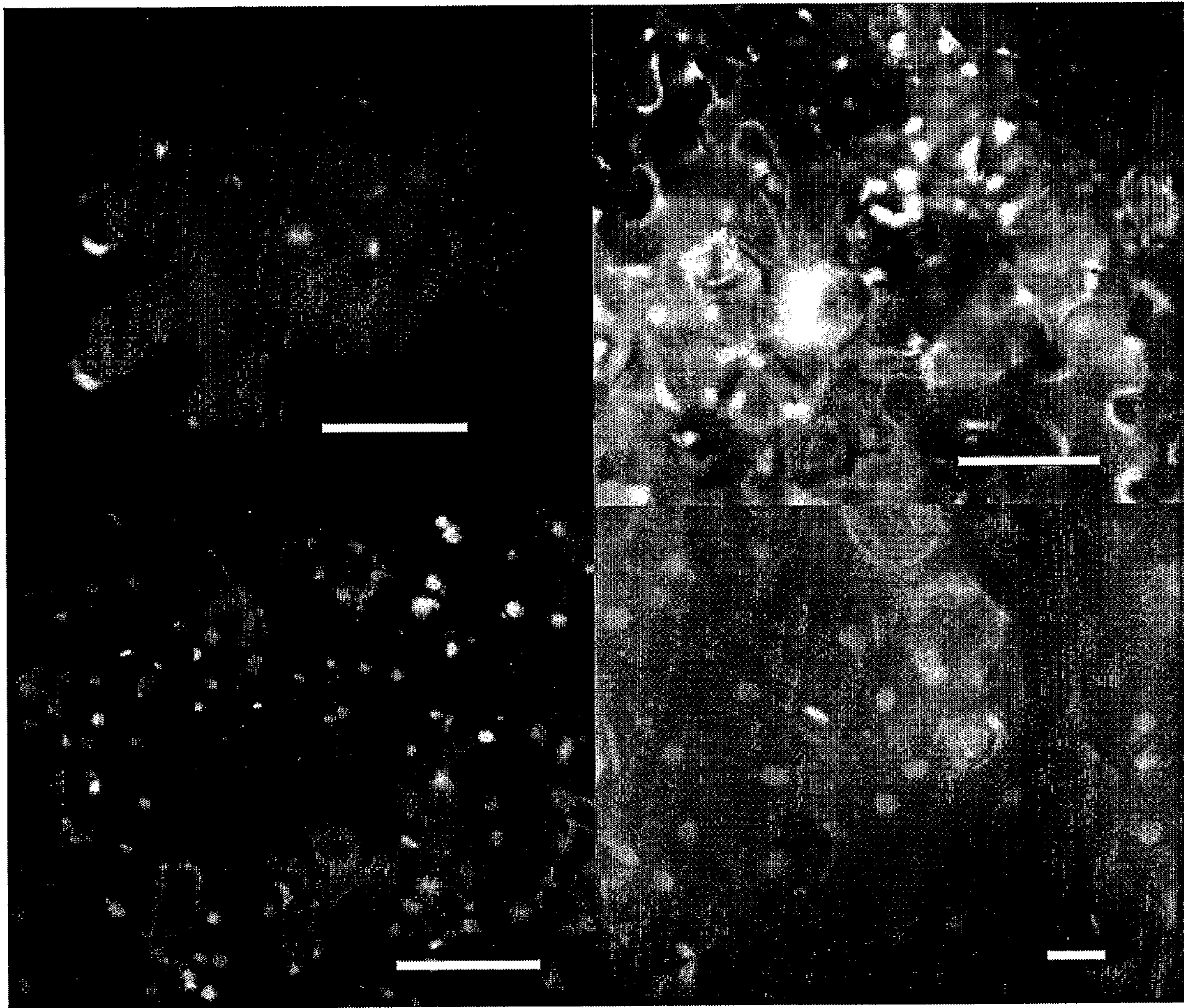


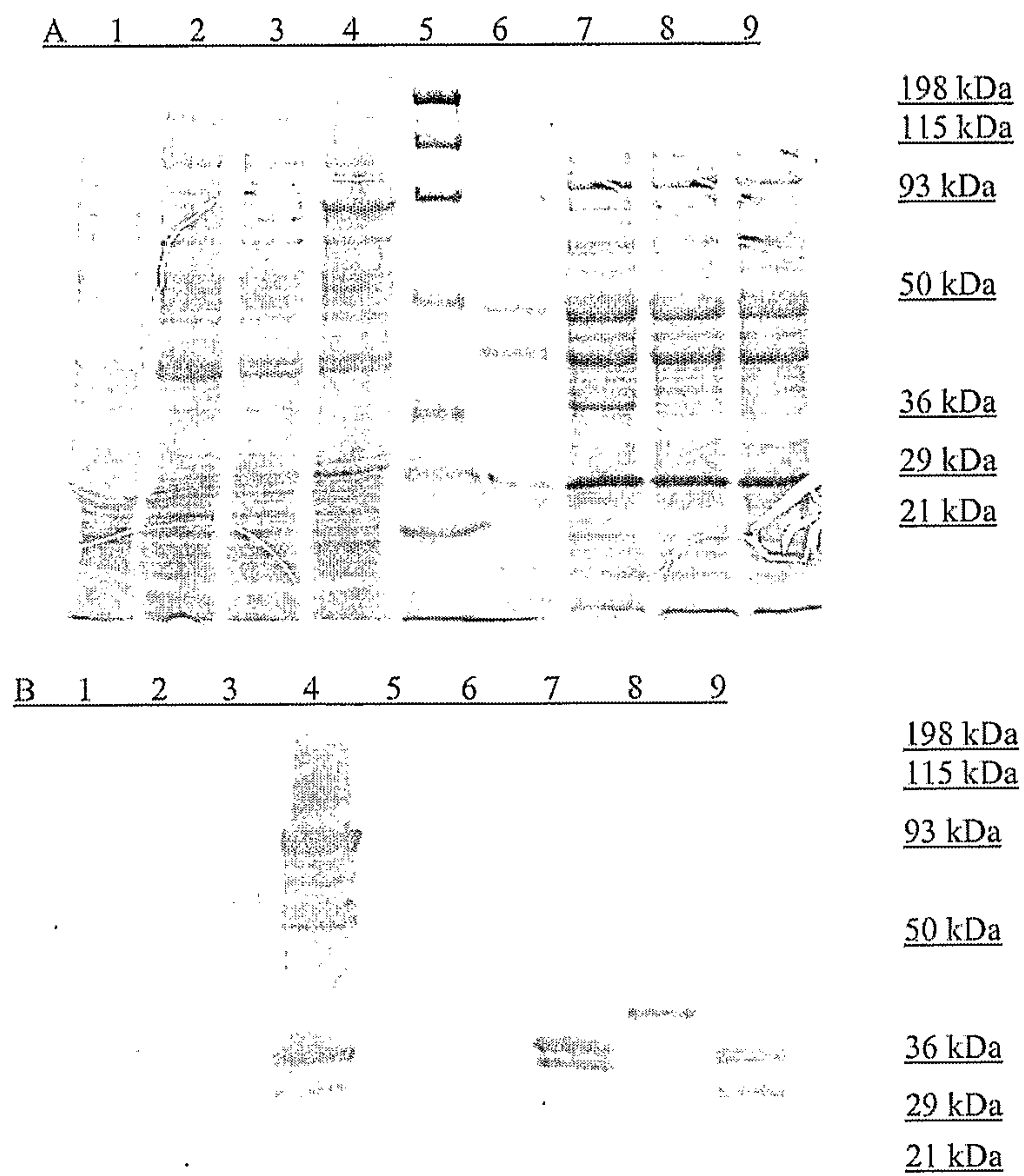
FIGURE 12

FIGURE 13

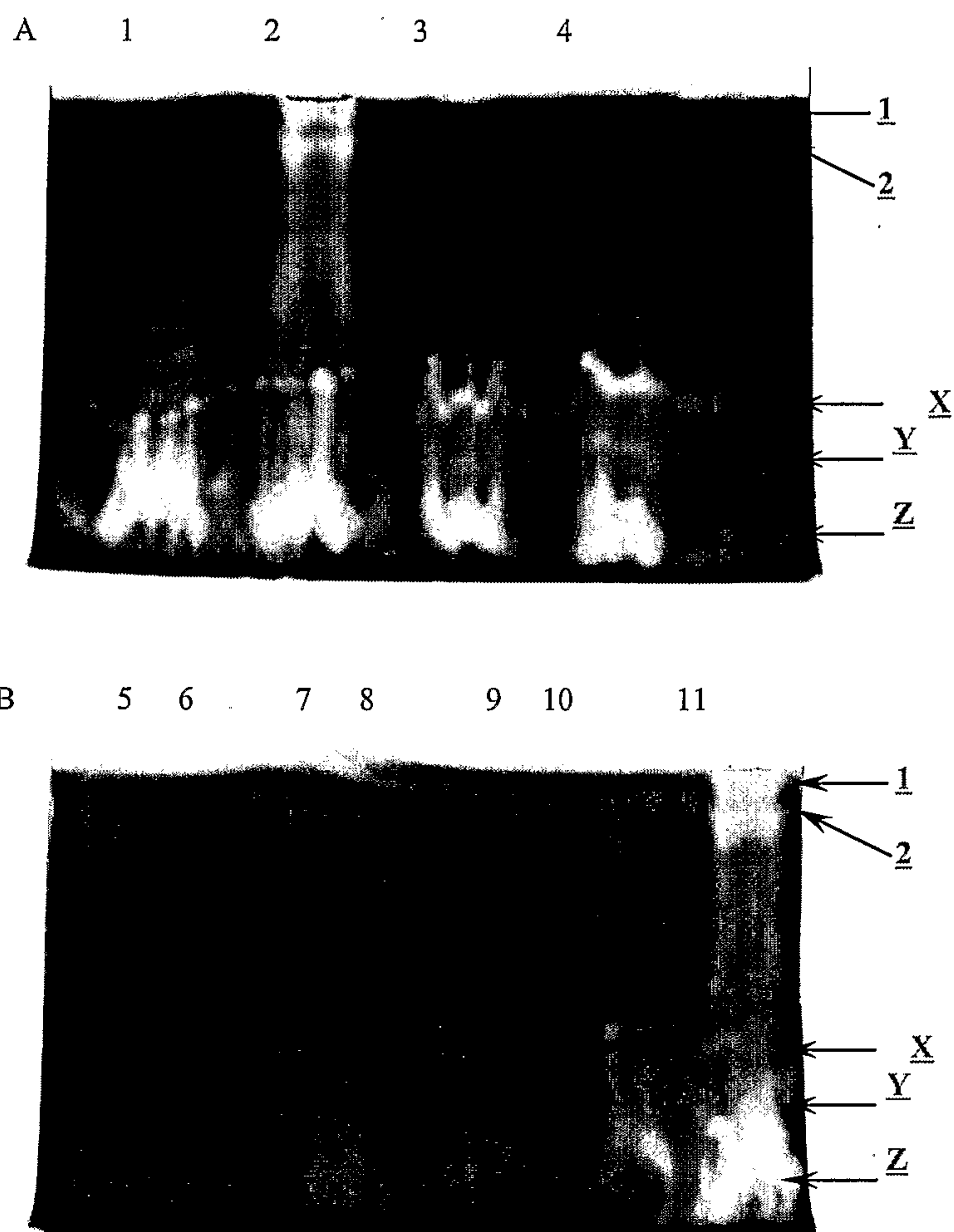


FIGURE 14

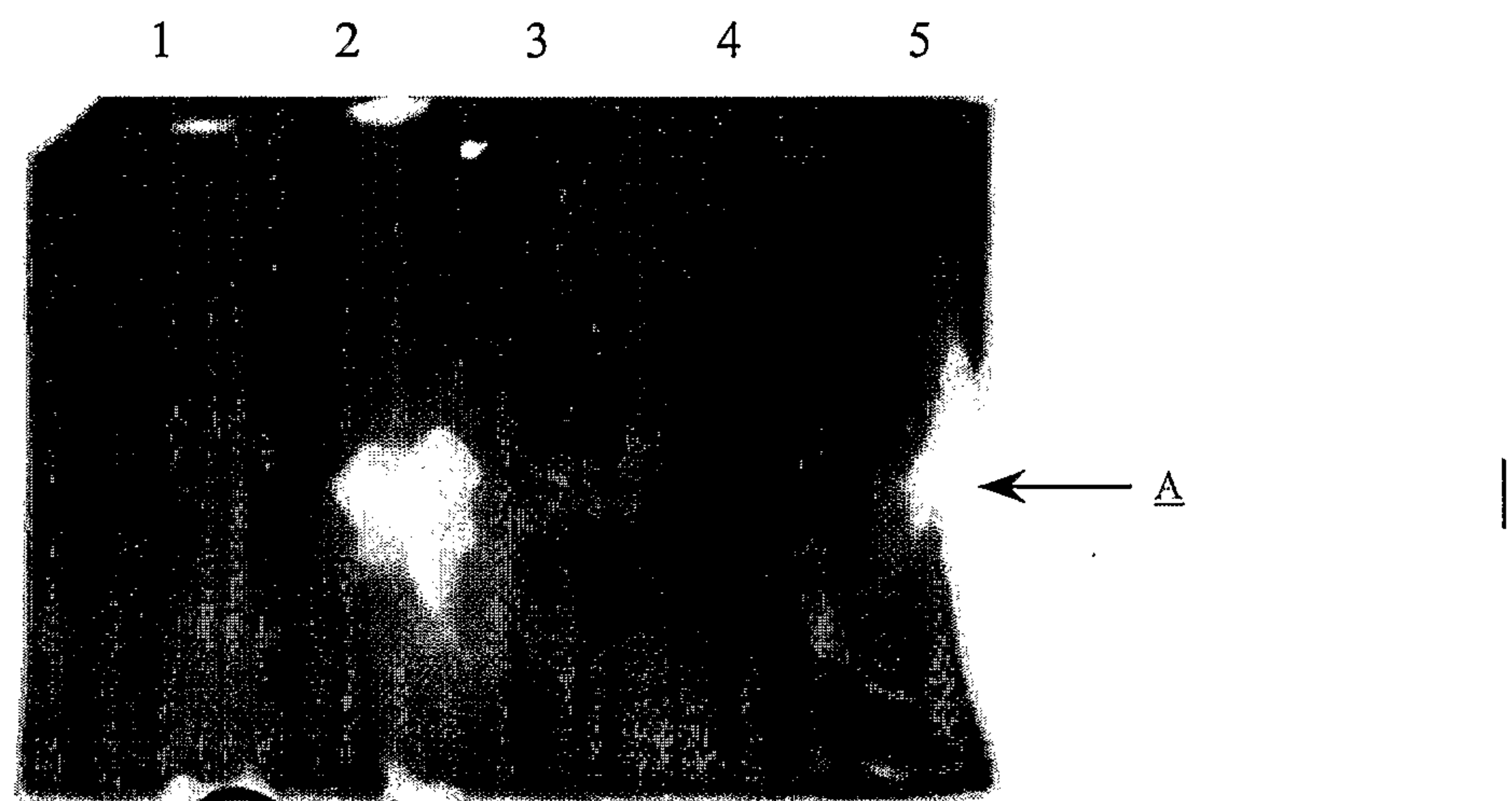


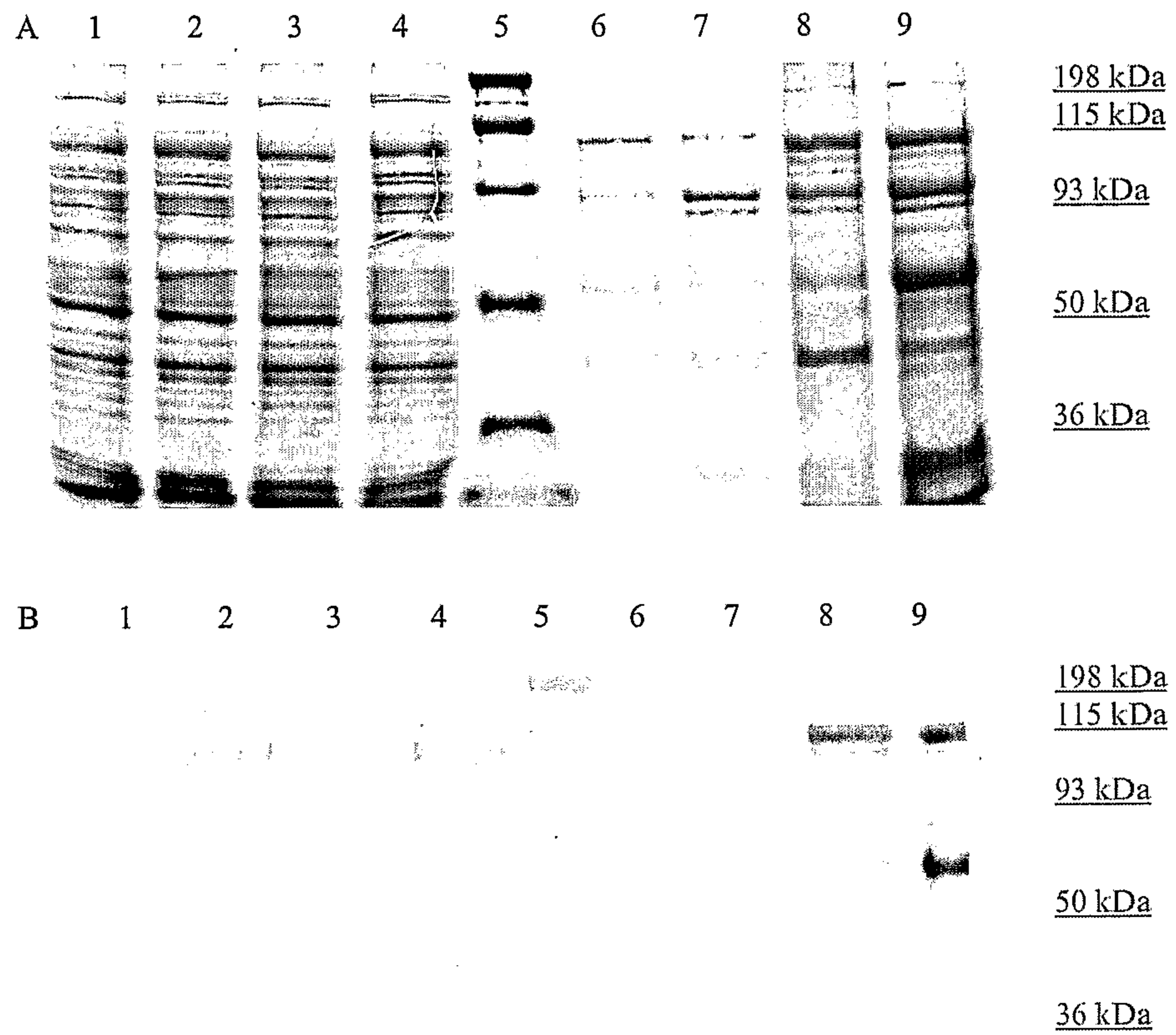
FIGURE 15

FIGURE 16