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

(19) World Intellectual Property Organization
International Bureau

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

PCT

(10) International Publication Number
WO 2006/018432 A2

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

(21) International Application Number:
PCT/EP2005/054034

(22) International Filing Date: 16 August 2005 (16.08.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

04103926.4	16 August 2004 (16.08.2004)	EP
60/602,680	19 August 2004 (19.08.2004)	US

(71) Applicant *(for all designated States except US)*: **CROPDESIGN N.V.** [BE/BE]; Technologiepark 3, B-9052 Zwijnaarde (BE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **FRANKARD, Valerie** [BE/BE]; Bierenberg 47, B-1640 Rhodes-Saint-Genèse (BE). **REUZEAU, Christophe** [FR/FR]; La Chapelle Gonaguet, F-24350 La Chapelle Gonaguet (FR). **SANZ MOLINERO, Ana, Isabel** [BE/BE]; Bernheimlaan 38, B-9050 Gentbrugge (BE).

(74) Common Representative: **CROPDESIGN N.V.**; Technologiepark 3, B-9052 Zwijnaarde (BE).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

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

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

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

WO 2006/018432 A2

(54) Title: PLANTS HAVING IMPROVED GROWTH CHARACTERISTICS AND METHOD FOR MAKING THE SAME

(57) Abstract: The present invention concerns a method for improving the growth characteristics of plants by increasing activity in a plant of an RNA-binding protein or a homologue thereof, wherein said RNA-binding protein or homologue thereof is either: (i) a polypeptide having RNA-binding activity and comprising either 2 or 3 RNA 10 recognition motifs (RRMs) and a motif having at least 75% sequence identity to motif I: PIYEAAVVALPVVVKERLVRILRLGIATRYD and/or a motif having at least 50% sequence identity to motif II: RFDPFTGEPYKFDP; or (ii) an RBP1 polypeptide or homologue thereof having (a) RNA-binding activity; (b) two RRM domains, (c) the following two motifs: (i) KIFVGGL; and (ii) 15 RPRGFGF, allowing for up to three amino acid substitutions and any conservative change in the motifs; and (d) having at least 20% sequence identity to the amino acid represented by SEQ ID NO: 15. The invention also concerns to transgenic plants having introduced therein an RNA-binding protein-encoding nucleic acid or variant thereof, which plants have improved growth characteristics relative to corresponding wild type plants. The present invention also concerns constructs useful in the methods of the invention.

Plants having improved growth characteristics and method for making the same

The present invention relates generally to the field of molecular biology and concerns a method for improving plant growth characteristics. More specifically, the present invention concerns a method for improving plant growth characteristics, in particular yield, by increasing activity in a plant of an RNA-binding protein or a homologue thereof. The present invention also concerns plants having increased activity of an RNA-binding protein or a homologue thereof, which plants have improved growth characteristics relative to corresponding wild type plants. The RNA-binding protein or homologue thereof useful in the methods of the invention is one having RNA binding activity and having either 2 or 3 RNA recognition motifs (RRMs) and which comprises a motif having at least 75% sequence identity to motif I: PYEAAVVALPVVVKERLVRILRLGIATRYD and/or a motif having at least 50% sequence identity to motif II: RFDPFTGEPYKFDP. The RNA-binding protein or homologue thereof useful in the methods of the invention may also be an RBP1 or homologue thereof having the following: (a) RNA-binding activity; (b) two RRM domains, (c) the following two motifs: (i) KIFVGGL; and (ii) RPRGFGF, allowing for up to three amino acid substitutions and any conservative change in the motifs; and (d) having, in increasing order of preference, at least 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% sequence identity to the amino acid represented by SEQ ID NO: 15. The invention also provides constructs useful in the methods of the invention.

The ever-increasing world population and the dwindling supply of arable land available for agriculture fuel agricultural research towards improving the efficiency of agriculture. Conventional means for crop and horticultural improvements utilise selective breeding techniques to identify plants having desirable characteristics. However, such selective breeding techniques have several drawbacks, namely that these techniques are typically labour intensive and result in plants that often contain heterogeneous genetic components that may not always result in the desirable trait being passed on from parent plants. Advances in molecular biology have allowed mankind to modify the germplasm of animals and plants. Genetic engineering of plants entails the isolation and manipulation of genetic material (typically in the form of DNA or RNA) and the subsequent introduction of that genetic material into a plant. Such technology has the capacity to deliver crops or plants having various improved economic, agronomic or horticultural traits. A trait of particular economic interest is yield. Yield is normally defined as the measurable produce of economic value from a crop. This may be defined in terms of quantity and/or quality. Yield is directly dependent on several factors, for example, the number and size of the organs, plant architecture (for example, the

number of branches), seed production and more. Root development, nutrient uptake and stress tolerance may also be important factors in determining yield. Crop yield may therefore be increased by optimizing one of the abovementioned factors.

The ability to improve various growth characteristics of a plant would have many applications in areas such as crop enhancement, plant breeding, in the production of ornamental plants, aboriculture, horticulture and forestry. Improving growth characteristics, such as yield may also find use in the production of algae for use in bioreactors (for the biotechnological production of substances such as pharmaceuticals, antibodies, or vaccines, or for the bioconversion of organic waste) and other such areas.

It has now been found that increasing activity in a plant of an RNA-binding protein or a homologue thereof gives plants having improved growth characteristics relative to corresponding wild type plants, which RNA-binding protein or homologue thereof has RNA binding activity and either 2 or 3 RNA recognition motifs (RRMs) and which comprises a motif having at least 75% sequence identity to motif I: PYEAAVVALPVVVKERLVRILRLGIATRYD and/or a motif having at least 50% sequence identity to motif II: RFDPFTGEPYKFDP. It has also now been found that increasing activity in a plant of an RBP1 polypeptide or homologue thereof gives plants having improved growth characteristics relative to corresponding wild type plants. The RBP1 or homologue thereof refers to a polypeptide having the following: (a) RNA-binding activity; (b) two RRM domains, (c) the following two motifs: (i) KIFVGGL; and (ii) RPRGFGF, allowing for up to three amino acid substitutions and any conservative change in the motifs; and (d) having, in increasing order of preference, at least 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% sequence identity to the amino acid represented by SEQ ID NO: 15.

RNA-binding proteins have an important role to play in the regulation of gene expression both at a transcriptional and posttranscriptional level. The level of regulation extends over all steps in the synthesis, processing and turnover of RNA molecules, including pre-mRNA splicing, polyadenylation, mRNA transport, translation and stability/decay. Regulation is mainly achieved either directly by RNA-binding proteins or indirectly, whereby RNA-binding proteins modulate the function of other regulatory factors. RNA-protein interactions are central to many aspects of cellular metabolism, cell differentiation and development, as well as to the replication of infectious pathogens. RNA recognition motifs or RRMs are typically present in a large variety of RNA-binding proteins and are involved in all post-transcriptional processes, whereby the number of RRMs per protein varies from one to four copies. The RRM is a region of around eighty amino acids containing several well conserved residues, some of which

cluster into two short submotifs, RNP-1 (octamer) and RNP-2 (hexamer) (Birney *et al.*, Nucleic Acids Research, 1993, Vol. 21, No. 25, 5803-5816).

The *Arabidopsis* genome encodes 196 RRM-containing proteins, an example of which is RBP1 (Lorkovic *et al.*, Nucleic Acids Research, 2002, Vol. 30, No. 3, 623-635). They report that the RRMs of *At*RBP1 are most similar to those of the metazoan Musashi proteins. In addition to *At*RBP1, Lorkovic *et al.* describe three proteins having similarity to *At*RBP1 and Musashi proteins. RBP1 from *Arabidopsis thaliana* was first isolated by Suzuki *et al.* (Plant Cell Physiol. 41(3): 282-288 (2000)) and was found to be expressed in rapidly dividing tissue. RBP1, an RNA-binding protein (as shown by Suzuki *et al.* 2000) comprises two RRMs.

According to one embodiment of the present invention, there is provided a method for improving the growth characteristics of a plant, comprising increasing activity in a plant of an RNA-binding protein or a homologue thereof, which RNA-binding protein or homologue thereof has RNA binding activity and either 2 or 3 RNA recognition motifs (RRMs) and which comprises a motif having at least 75% sequence identity to motif I: PYEAAVVALPVVVKERLVRILRLGIATRYD and/or a motif having at least 50% sequence identity to motif II: RFDPFTGEPYKFDP.

According to another embodiment of the present invention, there is provided a method for improving the growth characteristics of a plant, comprising increasing activity in a plant of an RBP1 polypeptide or a homologue thereof having the following: (a) RNA-binding activity; (b) two RRM domains, (c) the following two motifs: (i) KIFVGGL; and (ii) RPRGFGF, allowing for up to three amino acid substitutions and any conservative change in the motifs; and (d) having, in increasing order of preference, at least 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% sequence identity to the amino acid represented by SEQ ID NO: 15.

Advantageously, performance of the methods according to the present invention result in plants having a variety of improved growth characteristics, especially increased yield, particularly seed yield.

The term "increased yield" as defined herein is taken to mean an increase in any one or more of the following, each relative to corresponding wild type plants: (i) increased biomass (weight) of one or more parts of a plant, particularly aboveground (harvestable) parts, increased root biomass or increased biomass of any other harvestable part; (ii) increased seed yield, which includes an increase in seed biomass (seed weight) and which may be an increase in the seed

weight per plant or on an individual seed basis; (iii) increased number of (filled) seeds; (iv) increased seed size, which may also influence the composition of seeds; (v) increased seed volume, which may also influence the composition of seeds; (vi) increased harvest index, which is expressed as a ratio of the yield of harvestable parts, such as seeds, over the total biomass; and (vii) increased thousand kernel weight (TKW), which is extrapolated from the total weight of the number of filled seeds. An increased TKW may result from an increased seed size and/or seed weight.

Taking corn as an example, a yield increase may be manifested as one or more of the following: increase in the number of plants per hectare or acre, an increase in the number of ears per plant, an increase in the number of rows, number of kernels per row, kernel weight, thousand kernel weight, ear length/diameter, among others. Taking rice as an example, a yield increase may be manifested by an increase in one or more of the following: number of plants per hectare or acre, number of panicles per plant, number of spikelets per panicle, number of flowers per panicle, increase in the seed filling rate, increase in thousand kernel weight, among others. An increase in yield may also result in modified architecture, or may occur as a result of modified architecture.

According to a preferred feature, performance of the methods of the invention result in plants having increased yield. Therefore, according to the present invention, there is provided a method for increasing plant yield, which method comprises increasing activity in a plant of an RNA-binding protein or a homologue thereof, which RNA-binding protein or homologue thereof has RNA binding activity and either 2 or 3 RNA recognition motifs (RRMs) and which comprises a motif having at least 75% sequence identity to motif I: PYEAAVVALPVVVKERLVRILRLGIATRYD and/or a motif having at least 50% sequence identity to motif II: RFDPFTGEPYKFDP. According to another preferred feature of the present invention, there is provided a method for increasing plant yield, which method comprises increasing activity in a plant of an RBP1 polypeptide or a homologue thereof having the following: (a) RNA-binding activity; (b) two RRM domains, (c) the following two motifs: (i) KIFVGGL; and (ii) RPRGFGF, allowing for up to three amino acid substitutions and any conservative change in the motifs; and (d) having, in increasing order of preference, at least 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% sequence identity to the amino acid represented by SEQ ID NO: 15.

Since the transgenic plants according to the present invention have increased yield, it is likely that these plants exhibit an increased growth rate (during at least part of their life cycle), relative to the growth rate of corresponding wild type plants at a corresponding stage in their

life cycle. The increased growth rate may be specific to one or more parts of a plant (including seeds), or may be throughout substantially the whole plant. A plant having an increased growth rate may even exhibit early flowering. The increase in growth rate may take place at one or more stages in the life cycle of a plant or during substantially the whole plant life cycle. Increased growth rate during the early stages in the life cycle of a plant may reflect enhanced vigour. The increase in growth rate may alter the harvest cycle of a plant allowing plants to be sown later and/or harvested sooner than would otherwise be possible. If the growth rate is sufficiently increased, it may allow for the sowing of further seeds of the same plant species (for example sowing and harvesting of rice plants followed by sowing and harvesting of further rice plants all within one conventional growing period). Similarly, if the growth rate is sufficiently increased, it may allow for the sowing of further seeds of different plants species (for example the sowing and harvesting of rice plants followed by, for example, the sowing and optional harvesting of soy bean, potato or any other suitable plant). Harvesting additional times from the same rootstock in the case of some plants may also be possible. Altering the harvest cycle of a plant may lead to an increase in annual biomass production per acre (due to an increase in the number of times (say in a year) that any particular plant may be grown and harvested). An increase in growth rate may also allow for the cultivation of transgenic plants in a wider geographical area than their wild-type counterparts, since the territorial limitations for growing a crop are often determined by adverse environmental conditions either at the time of planting (early season) or at the time of harvesting (late season). Such adverse conditions may be avoided if the harvest cycle is shortened. The growth rate may be determined by deriving various parameters from growth curves plotting growth experiments, such parameters may be: T-Mid (the time taken for plants to reach 50% of their maximal size) and T-90 (time taken for plants to reach 90% of their maximal size), amongst others.

Performance of the methods of the invention gives plants having an increased growth rate. Therefore, according to the present invention, there is provided a method for increasing the growth rate of plants, which method comprises increasing activity in a plant of an RNA-binding protein or a homologue thereof, which RNA-binding protein or homologue thereof has RNA binding activity and either 2 or 3 RNA recognition motifs (RRMs) and which comprises a motif having at least 75% sequence identity to motif I: PYEAAVVALPVVVKERLVRILRLGIATRYD and/or a motif having at least 50% sequence identity to motif II: RFDPFTGEPYKFDP. There is also provided a further method for increasing the growth rate of plants, which method comprises increasing activity in a plant of an RBP1 polypeptide or a homologue thereof having the following: (a) RNA-binding activity; (b) two RRM domains, (c) the following two motifs: (i) KIFVGGL; and (ii) RPRGFGF, allowing for up to three amino acid substitutions and any conservative change in the motifs; and (d) having, in increasing order of preference, at least

20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% sequence identity to the amino acid represented by SEQ ID NO: 15.

An increase in yield and/or growth rate occurs whether the plant is under non-stress conditions or whether the plant is exposed to various mild stresses compared to control plants. Plants typically respond to exposure to stress by growing more slowly. In conditions of severe stress, the plant may even stop growing altogether. Mild stress on the other hand is defined herein as being any stress to which a plant is exposed which does not result in the plant ceasing to grow altogether without the capacity to resume growth. Due to advances in agricultural practices (irrigation, fertilization, pesticide treatments) severe stresses are not often encountered in cultivated crop plants. As a consequence, the compromised growth induced by mild stress is often an undesirable feature in agriculture. Mild stresses are the typical stresses to which a plant may be exposed. These stresses may be the everyday biotic and/or abiotic (environmental) stresses to which a plant is exposed. Typical abiotic or environmental stresses include temperature stresses caused by atypical hot or cold/freezing temperatures; salt stress; water stress (drought or excess water). Abiotic stresses may also be caused by chemicals. Biotic stresses are typically those stresses caused by pathogens, such as bacteria, viruses, fungi and insects.

The abovementioned growth characteristics may advantageously be modified in any plant.

The term "plant" as used herein encompasses whole plants, ancestors and progeny of plants and plant parts, including seeds, shoots, stems, leaves, roots, flowers (including tubers), and tissues and organs, wherein each of the aforementioned comprise the gene/nucleic acid of interest. The term "plant" also encompasses suspension cultures, callus tissue, embryos, meristematic regions, gametophytes, sporophytes, pollen, and microspores, again wherein each of the aforementioned comprise the gene/nucleic acid of interest.

Plants that are particularly useful in the methods of the invention include all plants which belong to the superfamily Viridiplantae, in particular monocotyledonous and dicotyledonous plants including fodder or forage legumes, ornamental plants, food crops, trees or shrubs selected from the list comprising *Acacia* spp.,*Acer* spp., *Actinidia* spp., *Aesculus* spp., *Agathis australis, Albizia amara, Alsophila tricolor, Andropogon* spp., *Arachis* spp, *Areca catechu, Astelia fragrans, Astragalus cicer, Baikiaea plurijuga, Betula* spp., *Brassica* spp., *Bruguiera gymnorrhiza, Burkea africana, Butea frondosa, Cadaba farinosa, Calliandra* spp, *Camellia sinensis, Canna indica, Capsicum* spp., *Cassia* spp., *Centroema pubescens, Chaenomeles* spp., *Cinnamomum cassia, Coffea arabica, Colophospermum mopane, Coronillia varia,*

Cotoneaster serotina, Crataegus spp., *Cucumis* spp., *Cupressus* spp., *Cyathea dealbata, Cydonia oblonga, Cryptomeria japonica, Cymbopogon* spp., *Cynthea dealbata, Cydonia oblonga, Dalbergia monetaria, Davallia divaricata, Desmodium* spp., *Dicksonia squarosa, Diheteropogon amplectens, Dioclea* spp, *Dolichos* spp., *Dorycnium rectum, Echinochloa pyramidalis, Ehrartia* spp., *Eleusine coracana, Eragrestis* spp., *Erythrina* spp., *Eucalyptus* spp., *Euclea schimperi, Eulalia villosa, Fagopyrum* spp., *Feijoa sellowiana, Fragaria* spp., *Flemingia* spp, *Freycinetia banksii, Geranium thunbergii, Ginkgo biloba, Glycine javanica, Gliricidia* spp, *Gossypium hirsutum, Grevillea* spp., *Guibourtia coleosperma, Hedysarum* spp., *Hemarthia altissima, Heteropogon contortus, Hordeum vulgare, Hyparrhenia rufa, Hypericum erectum, Hyperthelia dissoluta, Indigo incarnata, Iris* spp., *Leptarrhena pyrolifolia, Lespediza* spp., *Lettuca* spp., *Leucaena leucocephala, Loudetia simplex, Lotonus bainesii, Lotus* spp., *Macrotyloma axillare, Malus* spp., *Manihot esculenta, Medicago sativa, Metasequoia glyptostroboides, Musa sapientum, Nicotianum* spp., *Onobrychis* spp., *Ornithopus* spp., *Oryza* spp., *Peltophorum africanum, Pennisetum* spp., *Persea gratissima, Petunia* spp., *Phaseolus* spp., *Phoenix canariensis, Phormium cookianum, Photinia* spp., *Picea glauca, Pinus* spp., *Pisum sativum, Podocarpus totara, Pogonarthria fleckii, Pogonarthria squarrosa, Populus* spp., *Prosopis cineraria, Pseudotsuga menziesii, Pterolobium stellatum, Pyrus communis, Quercus* spp., *Rhaphiolepsis umbellata, Rhopalostylis sapida, Rhus natalensis, Ribes grossularia, Ribes* spp., *Robinia pseudoacacia, Rosa* spp., *Rubus* spp., *Salix* spp., *Schyzachyrium sanguineum, Sciadopitys verticillata, Sequoia sempervirens, Sequoiadendron giganteum, Sorghum bicolor, Spinacia* spp., *Sporobolus fimbriatus, Stiburus alopecuroides, Stylosanthos humilis, Tadehagi* spp, *Taxodium distichum, Themeda triandra, Trifolium* spp., *Triticum* spp., *Tsuga heterophylla, Vaccinium* spp., *Vicia* spp., *Vitis vinifera, Watsonia pyramidata, Zantedeschia aethiopica, Zea mays*, amaranth, artichoke, asparagus, broccoli, Brussels sprouts, cabbage, canola, carrot, cauliflower, celery, collard greens, flax, kale, lentil, oilseed rape, okra, onion, potato, rice, soybean, strawberry, sugar beet, sugarcane, sunflower, tomato, squash, tea and algae, amongst others. According to a preferred embodiment of the present invention, the plant is a crop plant such as soybean, sunflower, canola, alfalfa, rapeseed, cotton, tomato, potato or tobacco. Further preferably, the plant is a monocotyledonous plant, such as sugar cane. More preferably the plant is a cereal, such as rice, maize, wheat, barley, millet, rye, sorghum or oats.

The activity of an RNA-binding protein, or of a homologue thereof, may be increased by increasing levels of the RNA-binding protein. Alternatively, activity may also be increased without increase in levels of an RNA-binding protein, or even when there is a reduction in levels of an RNA-binding protein. This may occur when the intrinsic properties of the polypeptide are altered, for example, by making a mutant form that is more active that the wild

type. Similarly, the activity of an RBP1 polypeptide or homologue thereof may be increased by increasing levels of the RBP1 polypeptide protein. Alternatively, activity may also be increased when there is no change in levels of an RBP1, or even when there is a reduction in levels of an RBP1 polypeptide. This may occur when the intrinsic properties of the polypeptide are altered, for example, by making mutant that is more active that the wild type.

The term "RNA-binding protein or homologue thereof" as defined herein refers to a polypeptide with RNA binding activity and having either 2 or 3 RNA recognition motifs (RRMs) and which comprises a motif having at least 75%, 80%, 85%, 90% or 95% sequence identity to motif I: PYEAAVVALPVVVKERLVRILRLGIATRYD and/or a motif having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% or 95% sequence identity to motif II: RFDPFTGEPYKFDP. The term also refers to an amino acid sequence having in increasing order of preference at least 13%, 15%, 17%, 19%, 21%, 23%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% or 95% sequence identity to the amino acid sequence represented by SEQ ID NO: 2.

An "RNA-binding protein or a homologue thereof" falling within the above definition may readily be identified using routine techniques well known to persons skilled in the art. For example, RNA-binding activity may readily be determined *in vitro* or *in vivo* using techniques well known in the art. Examples of *in vitro* assays include: nucleic acid binding assays using North-Western and/or South-Western analysis (Suzuki *et al.* Plant Cell Physiol. 41(3): 282-288 (2000)); RNA binding assays using UV cross linking; Electrophoretic Mobility Shift Assay for RNA Binding Proteins (Smith, RNA-Protein Interactions - A Practical Approach 1998, University of Cambridge). Examples of *in vivo* assays include: TRAP (translational repression assay procedure) (Paraskeva E, Atzberger A, Hentze MW: A translational repression assay procedure (TRAP) for RNA-protein interactions *in vivo*. PNAS 1998 Feb 3; 95(3): 951-6.).

Whether a polypeptide has at least 13% identity to the amino acid represented by SEQ ID NO: 2 may readily be established by sequence alignment. Methods for the alignment of sequences for comparison are well known in the art, such methods include GAP, BESTFIT, BLAST, FASTA and TFASTA. GAP uses the algorithm of Needleman and Wunsch (J. Mol. Biol. 48: 443-453, 1970) to find the alignment of two complete sequences that maximises the number of matches and minimises the number of gaps. The BLAST algorithm calculates percent sequence identity and performs a statistical analysis of the similarity between the two sequences. The software for performing BLAST analysis is publicly available through the National Centre for Biotechnology Information. An RNA-binding protein or a homologue thereof having at least 13% identity to the amino acid represented by SEQ ID NO: 2 may

readily be identified by aligning a query sequence (preferably a protein sequence) with known RNA-binding protein sequences (see for example the alignment shown in Figure 1) using, for example, the VNTI AlignX multiple alignment program, based on a modified clustal W algorithm (InforMax, Bethesda, MD, http://www.informaxinc.com), with default settings for gap opening penalty of 10 and a gap extension of 0.05.

A person skilled in the art will also readily be able to identify motifs having at least 75%, 80%, 85%, 90% or 95% sequence identity to motif I: PYEAAVVALPVVVKERLVRILRLGIATRYD and/or motifs having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85% or 90% sequence identity to motif II: RFDPFTGEPYKFDP. This may easily be achieved by making an alignment and searching for homologous regions.

Table 1 below shows motif I and II as found in the sequence of SEQ ID NO: 2 and the percentage sequence identity with corresponding motifs in homologous RNA-binding proteins. RNA-binding proteins useful in the methods of the invention may contain motif I or II, or motifs I and II.

Table 1: Motifs found in RNA binding proteins and homologues thereof

	Gene name and Accession number	**Conserved Motif**	**% Sequence identity with the motifs SEQ ID NO: 2**
Motif I	**Tobacco CDS701 (SEQ ID NO: 2)**	**PYEAAVVALPVVVKERLVRILRLGIATRYD**	
	Rice CDS701 homologue (AL731884) SEQ ID NO: 4	PYEAAVVSLPSAVKELLLRILRLRIGTRYD	Identity: 23/30 (76.7%) # Similarity: 25/30 (83.3%)
	Rice predicted fragment AK059444 SEQ ID NO: 6	PYEAAVVSLPSAVKELLLRILRLRIGTRYD	Identity: 23/30 (76.7%) # Similarity: 25/30 (83.3%)
	Corn predicted fragment AY105295 SEQ ID NO: 8	PYESAVNSLPSAVKEVLLRILRLRIGTRYD	Identity: 21/30 (70.0%) # Similarity: 24/30 (80.0%)
	Consensus Motif I	PYE A/S AV V/N A/S LP V/S V/A VKE L/R/V L V/L RILRL G/R I A/G TRYD	30, 9 substitutions
Motif II	**Tobacco CDS701 (SEQ ID NO: 2)**	**RFDPFTGEPYKFDP**	

	Rice CDS701 homologue (AL731884) SEQ ID NO: 4	RFDPFTGEPYKFDP	Identity: 14/14 (100.0%) # Similarity: 14/14 (100.0%)
	Rice predicted fragment AK059444 SEQ ID NO: 6	RFDPFTGEPYKFDP	Identity: 14/14 (100.0%) # Similarity: 14/14 (100.0%)
	Corn predicted fragment AY105295 SEQ ID NO: 8	RFDPFTGEPYKFXP	Identity: 13/14 (92.9%) # Similarity: 13/14 (92.9%)
	Rice BAC83046 SEQ ID NO: 10	RYPPHLGEAIKFSP	Identity: 7/14 (50.0%) # Similarity: 8/14 (57.1%)
	Consensus M2	R F/Y D/P P F/H T/L GE P/A Y/I KF D/X/S	14, 7 substitutions

Examples of polypeptides falling under the definition of an "RNA-binding protein or a homologue thereof" include the following sequences: SEQ ID NO: 2 from tobacco; SEQ ID NO: 4 is a protein prediction of a BAC clone from rice (NCBI Accession number AL731884); SEQ ID NO: 6 is a rice protein prediction (fragment) from cDNA (NCBI Accession number AK059444); SEQ ID NO: 8 is a corn protein prediction (fragment) from cDNA (NCBI Accession number AY105295); and SEQ ID NO: 10 is a full length rice sequence (NCBI Accession number BAC83046).

It is to be understood that the term RNA-binding protein or a homologue thereof is not to be limited to the sequences represented by SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 and SEQ ID NO: 10, but that any polypeptide meeting the criteria of having RNA-binding activity and having either 2 or 3 RNA recognition motifs (RRMs) and which comprises a motif having at least 75% sequence identity to motif I: PYEAAVVALPVVVKERLVRILRLGIATRYD and/or a motif having at least 50% sequence identity to motif II: RFDPFTGEPYKFDP may also be useful in performing the methods of the invention.

The term "RBP1 or homologue thereof" as defined herein refers to a polypeptide having the following: (a) RNA-binding activity; (b) two RRM domains, (c) the following two motifs: (i) KIFVGGL; and (ii) RPRGFGF, allowing for up to three amino acid substitutions and any conservative change in the motifs; and (d) having, in increasing order of preference, at least 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%

sequence identity to the amino acid represented by SEQ ID NO: 15. Conservative substitution tables are well known in the art (see for example Creighton (1984) Proteins. W.H. Freeman and Company and see Table 4 below).

An "RBP1 polypeptide or a homologue thereof" falling within the above definition may readily be identified using routine techniques well known to persons skilled in the art. For example, RNA-binding activity may readily be determined as described above.

Furthermore, RRM domains are well known in the art and consist of around 80-90 amino acids; they have a structure consisting of four strands and two helices arranged in an alpha/beta sandwich, with a third helix sometimes being present during RNA binding. RRM domain-containing proteins have a modular structure. RRM domains may be identified using SMART (a Simple Modular Architecture Research Tool: Identification of signaling domains, Schultz *et al.* PNAS, 95, 5857-5864 (1998)) (http://smart.embl-heidelberg.de/). See also Letunic *et al.*, Recent improvements to the SMART domain-based sequence annotation resource (Nucleic Acids Res. 30(1), 242-244).

Whether a polypeptide has at least 20% identity to the amino acid represented by SEQ ID NO: 2 may readily be established by sequence alignment using the methods for alignment as described above.

Since RBP1 polypeptides comprise highly conserved regions, a person skilled in the art would readily be able to identify other RBP1 sequences by comparing any conserved regions of the query sequence against those of the known RBP1 sequences. Examples of these conserved regions include the following two motifs: (i) KIFVGGL; and (ii) RPRGFGF, allowing for up to three amino acid substitutions and any conservative change in the motifs.

Examples of polypeptides falling under the definition of an "RBP1 or a homologue thereof" include: At1g58470 (SEQ ID NO: 15), At4g26650 (SEQ ID NO: 17), At5g55550 (SEQ ID NO: 19), At4g14300 (SEQ ID NO: 21), At3g07810 (SEQ ID NO: 23), At2g33410 (SEQ ID NO: 25) and At5g47620 (SEQ ID NO: 27) all from *Arabidopsis thaliana*; NP_921939.1 (SEQ ID NO: 29) from rice; AK067725 (SEQ ID NO: 31) and AK070544 (SEQ ID NO: 33) which correspond to rice mRNAs encoding RBP1 polypeptides; CK210974 (SEQ ID NO: 35) from wheat and CA124210 (SEQ ID NO: 37) from sugarcane are partial protein predictions from ESTs (expressed sequence tags).

Despite what may appear to be a relatively low sequence homology (as low as approximately 25%), RPB1 proteins are highly conserved in structure with all full-length proteins having 2 RRM domains. rbp1 genes in other plant species may therefore easily be found (see the above examples from rice, sugarcane and wheat which have herein been identified for the first time as RBP1 proteins). Table 2 below shows the percentage identities for some of the sequences shown in the alignment of Fig. 3.

Table 2: Homology of RBP1 protein sequences with SEQ ID NO: 2 based on overall global sequence alignment

MIPs Accession Number Identifier (http://mips.gsf.de/)	SEQ ID NO	RRM domains	Global homology VNTI align program (informax)
At4g26650	SEQ ID NO: 17	2X RRM	28.4 %
At5g55550	SEQ ID NO: 19	2X RRM	28.9 %
At4g14300	SEQ ID NO: 21	2X RRM	31.9 %
At3g07810	SEQ ID NO: 23	2X RRM	24.9 %
At2g33410	SEQ ID NO: 25	2X RRM	29.2 %
At5g47620	SEQ ID NO: 27	2X RRM 2X RRM	26.7 %
AK070544-Os (DNA sequence corresponding to mRNA). Chromosomic location: BAC AC125782.2 (138541-142744)	SEQ ID NO: 33	2X RRM	26.8 %
AK067725-OS (DNA sequence corresponding to mRNA). Chromosomic location: BAC AP003747 (103016-107790)	SEQ ID NO: 31	2X RRM	26.3 %

It is to be understood that the term RBP1 polypeptide or a homologue thereof is not to be limited to the sequences represented by SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 23, SEQ ID NO: 25 or SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35 and SEQ ID NO: 37, but that any polypeptide meeting the criteria of having: (a) RNA-binding activity; (b) two RRM domains, (c) the following two motifs: (i) KIFVGGL; and (ii) RPRGFGF, allowing for up to three amino acid substitutions and any conservative change in the motifs; and (d) having, in increasing order of preference, at least 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% sequence identity to the amino acid represented by SEQ ID NO: 15 may be useful in performing the methods of the invention.

A nucleic acid encoding an RNA-binding protein or a homologue thereof may be any natural or synthetic nucleic acid. An RNA-binding protein or a homologue thereof as defined hereinabove is encoded by an RNA-binding protein-encoding nucleic acid/gene. Therefore the term "RNA-binding protein-encoding nucleic acid/gene" as defined herein is any nucleic acid/gene encoding an RNA-binding protein or a homologue thereof, as defined hereinabove. Examples of RNA-binding protein-encoding nucleic acids include those represented by any one of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7 and SEQ ID NO: 9. RNA-binding protein-encoding nucleic acids/genes and functional variants thereof may be suitable in practising the methods of the invention. Functional variant RNA-binding protein-encoding nucleic acid/genes include portions of an RNA-binding protein-encoding nucleic acid/gene and/or nucleic acids capable of hybridising with an RNA-binding protein-encoding nucleic acid/gene. The term "functional" in the context of a functional variant refers to a variant (i.e. a portion or a hybridising sequence) which encodes a polypeptide having RNA-binding activity and preferably and additionally at least one RRM, preferably either 2 or 3 RRMs and further preferably at least one of the following motifs: a motif having at least 75% sequence identity to motif I: PYEAAVVALPVVVKERLVRILRLGIATRYD and/or a motif having at least 50% sequence identity to motif II: RFDPFTGEPYKFDP. The term "functional may also refer to a nucleic acid encoding an RNA-binding protein or homologue thereof, as defined hereinabove, which when introduced and expressed in a plant gives plants having improved growth characteristics.

The nucleic acid encoding an RBP1 polypeptide or a homologue thereof may be any natural or synthetic nucleic acid. An RBP1 polypeptide or a homologue thereof as defined hereinabove is encoded by an rbp1 nucleic acid/gene. Therefore the term "rbp1 nucleic acid/gene" as defined herein is any nucleic acid/gene encoding an RBP1 polypeptide or a homologue thereof as defined hereinabove. Examples of rbp1 nucleic acids include those represented by any one of SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34 and SEQ ID NO: 36. rbp1 nucleic acids/genes and functional variants thereof may be suitable in practising the methods of the invention. Functional variant rbp1 nucleic acid/genes include portions of an rbp1 nucleic acid/gene and/or nucleic acids capable of hybridising with an rbp1 nucleic acid/gene. The term "functional" in the context of a functional variant refers to a variant (i.e. a portion or a hybridising sequence) which encodes a polypeptide having RNA-binding activity and at least one RRM domain, preferably two RRM domains and further preferably the following two motifs: (i) KIFVGGL and (ii) RPRGFGF, allowing for up to three amino acid substitutions and any conservative change in the motifs. The term "functional may also refer to

a nucleic acid encoding an RBP1 polypeptide or homologue thereof, as defined hereinabove, which when introduced and expressed in a plant gives plants having improved growth characteristics.

The term portion as defined herein refers to an RNA binding protein-encoding piece of DNA of, in increasing order of preference, at least 180, 300, 500 or 700 nucleotides in length and which portion encodes a polypeptide having RNA binding activity and at least 1 RRM, preferably two or three RRMs and at least one, preferably both, of motifs I or II. A portion may be prepared, for example, by making one or more deletions to an RNA-binding protein-encoding nucleic acid. The portions may be used in isolated form or they may be fused to other coding (or non coding) sequences in order to, for example, produce a protein that combines several activities, one of them being RNA binding activity. When fused to other coding sequences, the resulting polypeptide produced upon translation may be larger than that predicted for the RNA-binding protein portion. Preferably, the functional portion is a portion of a nucleic acid as represented by any one of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7 and SEQ ID NO: 9.

The term portion with reference to an rbp1 nucleic acid refers to a piece of DNA comprising at least 80 nucleotides and which portion encodes a polypeptide having RNA binding activity and having at least one RRM domain, preferably two RRM domains and further preferably the following two motifs: (i) KIFVGGL and (ii) RPRGFGF. A portion may be prepared, for example, by making one or more deletions to an rbp1 nucleic acid. The portions may be used in isolated form or they may be fused to other coding (or non coding) sequences in order to, for example, produce a protein that combines several activities, one of them being RNA binding activity. When fused to other coding sequences, the resulting polypeptide produced upon translation could be bigger than that predicted for the rbp1 fragment. Preferably, the functional portion is a portion of a nucleic acid as represented by any one of SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22 SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34 and SEQ ID NO: 36.

Another type of variant RNA-binding protein is a nucleic acid capable of hybridising under reduced stringency conditions, preferably under stringent conditions, with an RNA-binding protein-encoding nucleic acid/gene as hereinbefore defined, which hybridising sequence encodes a polypeptide having RNA binding activity and having at least 1 RRM, preferably two or three RRMs, and at least one, preferably two, of motifs I or II. The hybridising sequence is, in increasing order of preference, at least 180, 300, 500 or 700 nucleotides in length. Preferably, the hybridising sequence is capable of hybridising to a nucleic acid as represented by any one of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7 and SEQ ID NO: 9.

Similarly, another type of variant rbp1 is a nucleic acid capable of hybridising under reduced stringency conditions, preferably under stringent conditions, with an rbp1 nucleic acid/gene as hereinbefore defined, which hybridising sequence encodes a polypeptide having RNA binding activity and at least one RRM domain, preferably two RRM domains and further preferably the following two motifs: (i) KIFVGGL and (ii) RPRGFGF. The hybridising sequence is preferably at least 80 nucleotides in length. Preferably, the hybridising sequence is capable of hybridising to a nucleic acid as represented by any one of SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34 and SEQ ID NO: 36.

The term "hybridisation" as defined herein is a process wherein substantially homologous complementary nucleotide sequences anneal to each other. The hybridisation process can occur entirely in solution, i.e. where both complementary nucleic acids are in solution. The hybridisation process can also occur with one of the complementary nucleic acids immobilised to a matrix such as magnetic beads, Sepharose beads or any other resin. The hybridisation process can furthermore occur with one of the complementary nucleic acids immobilised to a solid support such as a nitro-cellulose or nylon membrane or immobilised by e.g. photolithography to, for example, a siliceous glass support (the latter known as nucleic acid arrays or microarrays or as nucleic acid chips). In order to allow hybridisation to occur, the nucleic acid molecules are generally thermally or chemically denatured to melt a double strand into two single strands and/or to remove hairpins or other secondary structures from single stranded nucleic acids. The stringency of hybridisation is influenced by conditions such as temperature, salt concentration, ionic strength and hybridisation buffer composition. Hybridisation occurs under reduced stringency conditions, preferably under stringent conditions. Examples of stringency conditions are shown in Table 3 below. Stringent conditions are those that are at least as stringent as, for example, conditions A-L; and reduced stringency conditions are at least as stringent as, for example, conditions M-R.

Table 3: Examples of stringency conditions

Stringency Condition	Polynucleotide Hybrid ±	Hybrid Length (bp) ‡	Hybridization Temperature and Buffer†	Wash Temperature and Buffer†
A	DNA:DNA	> or equal to 50	65 °C.; 1 ×SSC- or -42 °C.; 1 ×SSC, 50% formamide	65 °C.; 0.3 ×SSC

B	DNA:DNA	<50	Tb*; 1 ×SSC	Tb*; 1 ×SSC
C	DNA:RNA	> or equal to 50	67 °C.; 1 ×SSC- or -45 °C.; 1 ×SSC, 50% formamide	67 °C.; 0.3 ×SSC
D	DNA:RNA	<50	Td*; 1 ×SSC	Td*; 1 ×SSC
E	RNA:RNA	> or equal to 50	70 °C.; 1 ×SSC- or -50 °C.; 1 ×SSC, 50% formamide	70 °C.; 0.3 ×SSC
F	RNA:RNA	<50	Tf*; 1 ×SSC	Tf*; 1 ×SSC
G	DNA:DNA	> or equal to 50	65 °C.; 4 ×SSC- or -45 °C.; 4 ×SSC, 50% formamide	65 °C.; 1 ×SSC
H	DNA:DNA	<50	Th*; 4 °SSC	Th*; 4 ×SSC
I	DNA:RNA	> or equal to 50	67 °C.; 4 ×SSC- or -45 °C.; 4 ×SSC, 50% formamide	67 °C.; 1 ×SSC
J	DNA:RNA	<50	Tj*; 4 ×SSC	Tj*; 4 ×SSC
K	RNA:RNA	> or equal to 50	70 °C.; 4 ×SSC- or -40 °C.; 6 ×SSC, 50% formamide	67 °C.; 1 ×SSC
L	RNA:RNA	<50	Tl*; 2 ×SSC	Tl*; 2 ×SSC
M	DNA:DNA	> or equal to 50	50 °C.; 4 ×SSC- or -40 °C.; 6 ×SSC, 50% formamide	50 °C.; 2×SSC
N	DNA:DNA	<50	Tn*; 6 ×SSC	Tn*; 6 ×SSC
O	DNA:RNA	> or equal to 50	55 °C.; 4 ×SSC- or -42 °C.; 6 ×SSC, 50% formamide	55 ×C.; 2 ×SSC
P	DNA:RNA	<50	Tp*; 6 ×SSC	Tp*; 6 ×SSC
Q	RNA:RNA	> or equal to 50	60 °C.; 4 ×SSC- or -45°C.; 6 ×SSC, 50% formamide	60 °C.; 2 ×SSC
R	RNA:RNA	<50	Tr*; 4 ×SSC	Tr*; 4 ×SSC

‡ The "hybrid length" is the anticipated length for the hybridising nucleic acid. When nucleic acids of known sequence are hybridised, the hybrid length may be determined by aligning the sequences and identifying the conserved regions described herein.

† SSPE (1 ×SSPE is 0.15 M NaCl, 10 mM NaH2PO4, and 1.25 mM EDTA, pH 7.4) may be substituted for SSC (1 ×SSC is 0.15 M NaCl anmd 15 mM sodium citrate) in the hybridisation and wash buffers; washes are performed for 15 minutes after hybridisation is complete. The hybridisations and washes may additionally include 5 × Denhardt's reagent, .5-1.0% SDS, 100 ug/ml denatured, fragmented salmon # sperm DNA, 0.5% sodium pyrophosphate, and up to 50% formamide. *Tb-Tr: The hybridization temperature for hybrids anticipated to be less than 50 base pairs in length should be 5-10 °C less than the melting temperature Tm of the hybrids there Tm is determined according to the following equations. For hybrids less than 18 base pairs in length, Tm (°C.) = 2 (# of A + T bases) + 4 (# of G + C bases). For hybrids between 18 # and 49 base pairs in length, Tm (°C.) = 81.5 + 16.6 (log.sub.10[Na+]) + 0.41 (% G + C) - (600/N), where N is the number of bases in the hybrid, and [Na+] is the concentration of sodium ions in the hybridization buffer ([NA+] for 1 ×SSC = .165 M). ± The present invention encompasses the substitution of any one or more DNA or RNA hybrid partners with either a peptide nucleic acid (PNA) or a modified nucleic acid.

The RNA-binding protein-encoding nucleic acid or variant thereof may be derived from any natural or artificial source. The nucleic acid/gene or variant thereof may be isolated from a microbial source, such as bacteria, yeast or fungi, or from a plant, algae or animal (including human) source. This nucleic acid may be modified from its native form in composition and/or genomic environment through deliberate human manipulation. The nucleic acid is preferably of plant origin, whether from the same plant species (for example to the one in which it is to be introduced) or whether from a different plant species. The nucleic acid may be isolated from a dicotyledonous species, preferably from the family *Nicotianae*, further preferably from tobacco. More preferably, the RNA-binding protein-encoding nucleic acid isolated from tobacco is represented by SEQ ID NO: 1 and the RNA-binding protein amino acid sequence is as represented by SEQ ID NO: 2.

The rbp1 nucleic acid or variant thereof may be derived from any natural or artificial source. The nucleic acid/gene or variant thereof may be isolated from a microbial source, such as bacteria, yeast or fungi, or from a plant, algae or animal (including human) source. This nucleic acid may be modified from its native form in composition and/or genomic environment through deliberate human manipulation. The nucleic acid is preferably of plant origin, whether from the same plant species (for example to the one in which it is to be introduced) or whether

from a different plant species. The nucleic acid may be isolated from a dicotyledonous species, preferably from the family *Brassicaceae*, further preferably from *Arabidopsis thaliana*. More preferably, the rbp1 isolated from *Arabidopsis thaliana* is represented by SEQ ID NO: 14 and the RBP1 amino acid sequence is as represented by SEQ ID NO: 15.

The activity of an RNA-binding protein or a homologue thereof may be increased by introducing a genetic modification (preferably in the locus of an RNA-binding protein-encoding gene). Similarly, the activity of an RBP1 polypeptide or a homologue thereof may be increased by introducing a genetic modification (preferably in the locus of an rbp1 gene). The locus of a gene as defined herein is taken to mean a genomic region which includes the gene of interest and 10KB up- or downstream of the coding region.

The genetic modification may be introduced, for example, by any one (or more) of the following methods: TDNA activation, TILLING, site-directed mutagenesis, homologous recombination or by introducing and expressing in a plant a nucleic acid encoding an RNA-binding protein or a homologue thereof or by introducing and expressing in a plant a nucleic acid encoding an RBP1 polypeptide or a homologue thereof. Following introduction of the genetic modification there follows a step of selecting for increased activity of an RNA-binding protein or selecting for increased activity of an RBP1 polypeptide, which increase in activity gives plants having improved growth characteristics.

T-DNA activation tagging (Hayashi *et al.* Science (1992) 1350-1353) involves insertion of T-DNA usually containing a promoter (may also be a translation enhancer or an intron), in the genomic region of the gene of interest or 10KB up- or down stream of the coding region of a gene in a configuration such that the promoter directs expression of the targeted gene. Typically, regulation of expression of the targeted gene by its natural promoter is disrupted and the gene falls under the control of the newly introduced promoter. The promoter is typically embedded in a T-DNA. This T-DNA is randomly inserted into the plant genome, for example, through *Agrobacterium* infection and leads to overexpression of genes near to the inserted T-DNA. The resulting transgenic plants show dominant phenotypes due to overexpression of genes close to the introduced promoter. The promoter to be introduced may be any promoter capable of directing expression of a gene in the desired organism, in this case a plant. For example, constitutive, tissue-preferred, cell type-preferred and inducible promoters are all suitable for use in T-DNA activation.

A genetic modification may also be introduced in the locus of an RNA-binding protein-encoding gene using the technique of TILLING (Targeted Induced Local Lesions IN Genomes). This is a

mutagenesis technology useful to generate and/or identify, and to eventually isolate mutagenised variants of an RNA-binding protein-encoding nucleic acid (or rbp1-encoding nucleic acid) having RNA-binding protein activity. TILLING also allows selection of plants carrying such mutant variants. These mutant variants may even exhibit higher RNA-binding protein activity than that exhibited by the gene in its natural form. TILLING combines high-density mutagenesis with high-throughput screening methods. The steps typically followed in TILLING are: (a) EMS mutagenesis (Redei and Koncz, 1992; Feldmann *et al.*, 1994; Lightner and Caspar, 1998); (b) DNA preparation and pooling of individuals; (c) PCR amplification of a region of interest; (d) denaturation and annealing to allow formation of heteroduplexes; (e) DHPLC, where the presence of a heteroduplex in a pool is detected as an extra peak in the chromatogram; (f) identification of the mutant individual; and (g) sequencing of the mutant PCR product. Methods for TILLING are well known in the art (McCallum Nat Biotechnol. 2000 Apr; 18(4):455-7, reviewed by Stemple 2004 (TILLING--a high-throughput harvest for functional genomics. Nat Rev Genet. 2004 Feb;5(2):145-50.)).

Site directed mutagenesis may be used to generate variants of RNA-binding protein-encoding nucleic acids or portions thereof that retain activity, namely, RNA binding activity. Several methods are available to achieve site directed mutagenesis, the most common being PCR based methods (current protocols in molecular biology. Wiley Eds. http://www.4ulr.com/products/currentprotocols/index.html). Site directed mutagenesis may be used to generate variants of RNA-binding protein-encoding nucleic acids or portions thereof that retain activity, namely, RNA binding activity. Similarly, site directed mutagenesis may be used to generate variants of RBP1-encoding nucleic acids or portions thereof that retain activity, namely, RNA binding activity. Site directed mutagenesis may also be used to generate variants of RBP1-encoding nucleic acids or portions thereof that retain activity, namely, RNA binding activity.

TDNA activation, TILLING and site-directed mutagenesis are examples of technologies that enable the generation of novel alleles and RNA-binding protein variants that retain RNA-binding protein function or that enable the generation novel alleles and rbp1 variants that retain RBP1 function and which are therefore useful in the methods of the invention.

Homologous recombination allows introduction in a genome of a selected nucleic acid at a defined selected position. Homologous recombination is a standard technology used routinely in biological sciences for lower organisms such as yeast or moss (e.g. physcomitrella). Methods for performing homologous recombination in plants have been described not only for model plants (Offringa *et al.* Extrachromosomal homologous recombination and gene targeting

in plant cells after *Agrobacterium*-mediated transformation. 1990 EMBO J. 1990 Oct; 9(10):3077-84) but also for crop plants, for example rice (Terada R, Urawa H, Inagaki Y, Tsugane K, Iida S. Efficient gene targeting by homologous recombination in rice. Nat Biotechnol. 2002. Iida and Terada: A tale of two integrations, transgene and T-DNA: gene targeting by homologous recombination in rice. Curr Opin Biotechnol. 2004 Apr; 15(2):132-8). The nucleic acid to be targeted (which may be an RNA-binding protein-encoding nucleic acid or variant thereof as hereinbefore defined or which may be an rbp1 nucleic acid or variant thereof as hereinbefore defined) need not be targeted to the locus of an RNA-binding protein gene or targeted to the locus of an rbp1 gene, but may be introduced in, for example, regions of high expression. The nucleic acid to be targeted may be an improved allele used to replace the endogenous gene or may be introduced in addition to the endogenous gene.

According to a preferred embodiment of the invention, plant growth characteristics may be improved by introducing and expressing in a plant a nucleic acid encoding an RNA-binding polypeptide or a homologue thereof, which has RNA binding activity and either 2 or 3 RNA recognition motifs (RRMs) and which comprises a motif having at least 75% sequence identity to motif I: PYEAAVVALPVVVKERLVRILRLGIATRYD and/or a motif having at least 50% sequence identity to motif II: RFDPFTGEPYKFDP.

A preferred method for introducing a genetic modification (which in this case need not be in the locus of an RNA-binding protein gene) is to introduce and express in a plant a nucleic acid encoding an RNA-binding protein or a homologue thereof, as defined hereinabove.

According to a further preferred embodiment of the invention, plant growth characteristics may be improved by introducing and expressing in a plant a nucleic acid encoding an RBP1 polypeptide or a homologue thereof.

One preferred method for introducing a genetic modification (which in this case need not be in the locus of an rbp1 gene) is to introduce and express in a plant a nucleic acid encoding an RBP1 polypeptide or a homologue thereof. An RBP1 polypeptide or a homologue thereof as mentioned above is one having: (a) RNA-binding activity; (b) two RRM domains, (c) the following two motifs: (i) KIFVGGL; and (ii) RPRGFGF, allowing for up to three amino acid substitutions and any conservative change in the motifs; and (d) having, in increasing order of preference, at least 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% sequence identity to the amino acid represented by SEQ ID NO: 15.

"Homologues" of a protein encompass peptides, oligopeptides, polypeptides, proteins and enzymes having amino acid substitutions, deletions and/or insertions relative to the unmodified protein in question and having similar biological and functional activity as the unmodified protein from which they are derived. To produce such homologues, amino acids of the protein may be replaced by other amino acids having similar properties (such as similar hydrophobicity, hydrophilicity, antigenicity, propensity to form or break α-helical structures or β-sheet structures). Conservative substitution tables are well known in the art (see for example Creighton (1984) Proteins. W.H. Freeman and Company). The table below gives examples of conserved amino acid substitutions.

Table 4: Examples of conserved amino acid substitutions

Residue	Conservative Substitutions	Residue	Conservative Substitutions
Ala	Ser	Leu	Ile; Val
Arg	Lys	Lys	Arg; Gln
Asn	Gln; His	Met	Leu; Ile
Asp	Glu	Phe	Met; Leu; Tyr
Gln	Asn	Ser	Thr; Gly
Cys	Ser	Thr	Ser; Val
Glu	Asp	Trp	Tyr
Gly	Pro	Tyr	Trp; Phe
His	Asn; Gln	Val	Ile; Leu
Ile	Leu, Val		

Also encompassed by the term "homologues" are two special forms of homology, which include orthologous sequences and paralogous sequences, which encompass evolutionary concepts used to describe ancestral relationships of genes. The term "paralogous" relates to gene-duplications within the genome of a species leading to paralogous genes. The term "orthologous" relates to homologous genes in different organisms due to speciation.

Othologues in, for example, monocot plant species may easily be found by performing a so-called reciprocal blast search. This may be done by a first blast involving blasting the sequence in question (for example, SEQ ID NO: 1 or 2 or SEQ ID NO: 14 or 15) against any sequence database, such as the publicly available NCBI database which may be found at: http://www.ncbi.nlm.nih.gov. If orthologues in rice were sought, the sequence in question would be blasted against, for example, the 28,469 full-length cDNA clones from *Oryza sativa* Nipponbare available at NCBI. BLASTn or tBLASTX may be used when starting from nucleotides or BLASTP or TBLASTN when starting from the protein, with standard default

values. The blast results may be filtered. The full-length sequences of either the filtered results or the non-filtered results are then blasted back (second blast) against the sequences of the organism from which the sequence in question is derived. The results of the first and second blasts are then compared. An orthologue is found when the results of the second blast give as hits with the highest similarity an RNA-binding protein-encoding nucleic acid or RNA-binding protein polypeptide, for example, if one of the organisms is tobacco then a paralogue is found. For RBP1, an orthologue is found when the results of the second blast give as hits with the highest similarity an rbp1 nucleic acid or RBP1 polypeptide, for example, if one of the organisms is *Arabidopsis thaliana* then a paralogue is found. In the case of large families, ClustalW may be used, followed by a neighbour joining tree, to help visualize the clustering.

A homologue may be in the form of a "substitutional variant" of a protein, i.e. where at least one residue in an amino acid sequence has been removed and a different residue inserted in its place. Amino acid substitutions are typically of single residues, but may be clustered depending upon functional constraints placed upon the polypeptide; insertions will usually be of the order of about 1 to 10 amino acid residues. Preferably, amino acid substitutions comprise conservative amino acid substitutions.

A homologue may also be in the form of an "insertional variant" of a protein, i.e. where one or more amino acid residues are introduced into a predetermined site in a protein. Insertions may comprise amino-terminal and/or carboxy-terminal fusions as well as intra-sequence insertions of single or multiple amino acids. Generally, insertions within the amino acid sequence will be smaller than amino- or carboxy-terminal fusions, of the order of about 1 to 10 residues. Examples of amino- or carboxy-terminal fusion proteins or peptides include the binding domain or activation domain of a transcriptional activator as used in the yeast two-hybrid system, phage coat proteins, (histidine)6-tag, glutathione S-transferase-tag, protein A, maltose-binding protein, dihydrofolate reductase, Tag·100 epitope, c-myc epitope, FLAG®-epitope, lacZ, CMP (calmodulin-binding peptide), HA epitope, protein C epitope and VSV epitope.

Homologues in the form of "deletion variants" of a protein are characterised by the removal of one or more amino acids from a protein.

Amino acid variants of a protein may readily be made using peptide synthetic techniques well known in the art, such as solid phase peptide synthesis and the like, or by recombinant DNA manipulations. Methods for the manipulation of DNA sequences to produce substitution, insertion or deletion variants of a protein are well known in the art. For example, techniques for making substitution mutations at predetermined sites in DNA are well known to those

skilled in the art and include M13 mutagenesis, T7-Gen *in vitro* mutagenesis (USB, Cleveland, OH), QuickChange Site Directed mutagenesis (Stratagene, San Diego, CA), PCR-mediated site-directed mutagenesis or other site-directed mutagenesis protocols.

The RNA-binding protein or homologue thereof may be a derivative or the RBP1 polypeptide or homologue thereof may be a derivative. "Derivatives" include peptides, oligopeptides, polypeptides, proteins and enzymes which may comprise substitutions, deletions or additions of naturally and non-naturally occurring amino acid residues compared to the amino acid sequence of a naturally-occurring form of the protein, for example, as presented in SEQ ID NO: 2, or SEQ ID NO: 15 in the case of RBP1. "Derivatives" of a protein encompass peptides, oligopeptides, polypeptides, proteins and enzymes which may comprise naturally occurring altered, glycosylated, acylated or non-naturally occurring amino acid residues compared to the amino acid sequence of a naturally-occurring form of the polypeptide. A derivative may also comprise one or more non-amino acid substituents compared to the amino acid sequence from which it is derived, for example a reporter molecule or other ligand, covalently or non-covalently bound to the amino acid sequence, such as a reporter molecule which is bound to facilitate its detection, and non-naturally occurring amino acid residues relative to the amino acid sequence of a naturally-occurring protein.

The RNA-binding protein or homologue thereof may be encoded by an alternative splice variant of an RNA-binding protein nucleic acid/gene. The RBP1 polypeptide or homologue thereof may be encoded by an alternative splice variant of an rbp1 nucleic acid/gene. The term "alternative splice variant" as used herein encompasses variants of a nucleic acid sequence in which selected introns and/or exons have been excised, replaced or added. Such variants will be ones in which the biological activity of the protein is retained, which may be achieved by selectively retaining functional segments of the protein. Such splice variants may be found in nature or may be manmade. Methods for making such splice variants are well known in the art. Preferred splice variants are splice variants of the nucleic acid represented by SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7 and SEQ ID NO: 9. Further preferred are splice variants encoding a polypeptide retaining RNA-binding activity and having at least 1 RRM, preferably two or three RRMs and at least one, preferably both, of motifs I or II. Preferred splice variants of RBP1 are splice variants of the nucleic acid represented by SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 23, SEQ ID NO: 25 or SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35 and SEQ ID NO: 37. Further preferred are splice variants encoding a polypeptide retaining RNA-binding activity and having one, preferably two RRM domains and further preferably the

following two motifs: (i) KIFVGGL; and (ii) RPRGFGF, allowing for up to three amino acid substitutions and any conservative change in the motifs.

The homologue may also be encoded by an allelic variant of a nucleic acid encoding an RNA-binding protein or a homologue thereof, preferably an allelic variant of the nucleic acid represented by any one of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7 and SEQ ID NO: 9. Further preferably, the polypeptide encoded by the allelic variant has RNA-binding activity and at least 1 RRM, preferably two or three RRMs and at least one, preferably both, of motifs I or II. The homologue may also be encoded by an allelic variant of a nucleic acid encoding an RBP1 polypeptide or a homologue thereof, preferably an allelic variant of the nucleic acid represented by SEQ ID NO: SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 23, SEQ ID NO: 25 or SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35 and SEQ ID NO: 37. Further preferably, the polypeptide encoded by the allelic variant has RNA-binding activity and one, preferably two RRM domains and the following two motifs: (i) KIFVGGL; and (ii) RPRGFGF, allowing for up to three amino acid substitutions and any conservative change in the motifs. Allelic variants exist in nature and encompassed within the methods of the present invention is the use of these natural alleles. Allelic variants encompass Single Nucleotide Polymorphisms (SNPs), as well as Small Insertion/Deletion Polymorphisms (INDELs). The size of INDELs is usually less than 100 bp. SNPs and INDELs form the largest set of sequence variants in naturally occurring polymorphic strains of most organisms.

According to a preferred aspect of the present invention, enhanced or increased expression of the RNA-binding protein encoding nucleic acid or variant thereof is envisaged. According to a preferred aspect of the present invention, enhanced or increased expression of the rbp1 nucleic acid or variant thereof is envisaged. Methods for obtaining enhanced or increased expression of genes or gene products are well documented in the art and include, for example, overexpression driven by appropriate promoters, the use of transcription enhancers or translation enhancers. Isolated nucleic acids which serve as promoter or enhancer elements may be introduced in an appropriate position (typically upstream) of a non-heterologous form of a polynucleotide so as to upregulate expression of an RNA-binding protein-encoding nucleic acid or variant thereof. For example, endogenous promoters may be altered *in vivo* by mutation, deletion, and/or substitution (see, Kmiec, U.S. Pat. No. 5,565,350; Zarling *et al.*, PCT/US93/03868), or isolated promoters may be introduced into a plant cell in the proper orientation and distance from a gene of the present invention so as to control the expression of the gene.

If polypeptide expression is desired, it is generally desirable to include a polyadenylation region at the 3'-end of a polynucleotide coding region. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA. The 3' end sequence to be added may be derived from, for example, the nopaline synthase or octopine synthase genes, or alternatively from another plant gene, or less preferably from any other eukaryotic gene.

An intron sequence may also be added to the 5' untranslated region or the coding sequence of the partial coding sequence to increase the amount of the mature message that accumulates in the cytosol. Inclusion of a spliceable intron in the transcription unit in both plant and animal expression constructs has been shown to increase gene expression at both the mRNA and protein levels up to 1000-fold, Buchman and Berg, Mol. Cell biol. 8:4395-4405 (1988); Callis *et al.*, Genes Dev. 1:1183-1200 (1987). Such intron enhancement of gene expression is typically greatest when placed near the 5' end of the transcription unit. Use of the maize introns Adh1-S intron 1, 2, and 6, the Bronze-1 intron are known in the art. See generally, The Maize Handbook, Chapter 116, Freeling and Walbot, Eds., Springer, N.Y. (1994).

The invention also provides genetic constructs and vectors to facilitate introduction and/or expression of the nucleotide sequences useful in the methods according to the invention.

Therefore, there is provided a gene construct comprising:

(i) An RNA-binding protein-encoding nucleic acid or variant thereof;

(ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (i); and optionally

(iii) a transcription termination sequence.

There is also provided, a gene construct comprising:

(i) An rbp1 nucleic acid or variant thereof;

(ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (i); and optionally

(iii) a transcription termination sequence.

Constructs useful in the methods according to the present invention may be constructed using recombinant DNA technology well known to persons skilled in the art. The gene constructs may be inserted into (commercially available) vectors suitable for transforming into plants cells and suitable for expression of the gene of interest in the transformed cells.

Plants are transformed with a vector comprising the sequence of interest (i.e., an RNA-binding protein-encoding nucleic acid or variant thereof or an rbp1 nucleic acid or variant thereof). The sequence of interest is operably linked to one or more control sequences (at least to a promoter). The terms "regulatory element", "control sequence" and "promoter" are all used interchangeably herein and are to be taken in a broad context to refer to regulatory nucleic acid sequences capable of effecting expression of the sequences to which they are ligated. Encompassed by the aforementioned terms are transcriptional regulatory sequences derived from a classical eukaryotic genomic gene (including the TATA box which is required for accurate transcription initiation, with or without a CCAAT box sequence) and additional regulatory elements (i.e. upstream activating sequences, enhancers and silencers) which alter gene expression in response to developmental and/or external stimuli, or in a tissue-specific manner. Also included within the term is a transcriptional regulatory sequence of a classical prokaryotic gene, in which case it may include a –35 box sequence and/or –10 box transcriptional regulatory sequences. The term "regulatory element" also encompasses a synthetic fusion molecule or derivative which confers, activates or enhances expression of a nucleic acid molecule in a cell, tissue or organ. The term "operably linked" as used herein refers to a functional linkage between the promoter sequence and the gene of interest, such that the promoter sequence is able to initiate transcription of the gene of interest.

Advantageously, any type of promoter may be used to drive expression of the nucleic acid sequence. The promoter may be an inducible promoter, i.e. having induced or increased transcription initiation in response to a developmental, chemical, environmental or physical stimulus. An example of an inducible promoter being a stress-inducible promoter, i.e. a promoter activated when a plant is exposed to various stress conditions. Additionally or alternatively, the promoter may be a tissue-preferred promoter, i.e. one that is capable of predominantly initiating transcription in certain tissues, such as the leaves, roots, seed tissue etc.

Preferably, the RNA-binding protein-encoding nucleic acid or variant thereof is operably linked to a seed-preferred promoter. A seed-preferred promoter is one that preferentially, but not necessarily exclusively, drives expression in seed-tissue. Preferably, the seed-tissue is the endosperm. Preferably, the promoter is a prolamin promoter, such as the prolamin promoter from rice (SEQ ID NO: 11). It should be clear that the applicability of the present invention is not restricted to the RNA-binding protein-encoding nucleic acid represented by SEQ ID NO: 1, nor is the applicability of the invention restricted to expression of an RNA-binding protein-encoding nucleic acid when driven by a prolamin promoter.

Preferably, the rbp1 nucleic acid or variant thereof is operably linked to a promoter capable of preferentially expressing the nucleic acid in shoots. Preferably, the promoter capable of preferentially expressing the nucleic acid in shoots has a comparable expression profile to a beta-expansin promoter, for example as shown in Fig. 5. Most preferably, the promoter capable of preferentially expressing the nucleic acid in shoots is the beta-expansin promoter from rice (SEQ ID NO: 38). It should be clear that the applicability of the present invention is not restricted to the rbp1 nucleic acid represented by SEQ ID NO: 14, nor is the applicability of the invention restricted to expression of an rbp1 nucleic acid when driven by a beta expansin promoter.

Optionally, one or more terminator sequences may also be used in the construct introduced into a plant. The term "terminator" encompasses a control sequence which is a DNA sequence at the end of a transcriptional unit which signals 3' processing and polyadenylation of a primary transcript and termination of transcription. Additional regulatory elements may include transcriptional as well as translational enhancers. Those skilled in the art will be aware of terminator and enhancer sequences which may be suitable for use in performing the invention

The genetic constructs of the invention may further include an origin of replication sequence which is required for maintenance and/or replication in a specific cell type. One example is when a genetic construct is required to be maintained in a bacterial cell as an episomal genetic element (e.g. plasmid or cosmid molecule). Preferred origins of replication include, but are not limited to, the f1-ori and colE1.

The genetic construct may optionally comprise a selectable marker gene. As used herein, the term "selectable marker gene" includes any gene that confers a phenotype on a cell in which it is expressed to facilitate the identification and/or selection of cells that are transfected or transformed with a nucleic acid construct of the invention. Suitable markers may be selected from markers that confer antibiotic or herbicide resistance, that introduce a new metabolic trait or that allow visual selection. Examples of selectable marker genes include genes conferring resistance to antibiotics (such as *nptII* that phosphorylates neomycin and kanamycin, or *hpt*, phosphorylating hygromycin), to herbicides (for example *bar* which provides resistance to Basta; *aroA* or *gox* providing resistance against glyphosate), or genes that provide a metabolic trait (such as *manA* that allows plants to use mannose as sole carbon source). Visual marker genes result in the formation of colour (for example β-glucuronidase, GUS), luminescence (such as luciferase) or fluorescence (Green Fluorescent Protein, GFP, and derivatives thereof).

The present invention also encompasses plants obtainable by the methods according to the present invention. The present invention therefore provides plants obtainable by the method according to the present invention, which plants have introduced therein an RNA-binding protein-encoding nucleic acid or variant thereof or an rbp1 nucleic acid or variant thereof.

The invention also provides a method for the production of transgenic plants having improved growth characteristics, comprising introduction and expression in a plant of an RNA-binding protein-encoding nucleic acid or a variant thereof.

More specifically, the present invention provides a method for the production of transgenic plants having improved growth characteristics, which method comprises:

(i) introducing into a plant or plant cell an RNA-binding protein-encoding nucleic acid or variant thereof; and

(ii) cultivating the plant cell under conditions promoting plant growth and development.

The invention also provides a method for the production of transgenic plants having improved growth characteristics, comprising introduction and expression in a plant of an rbp1 nucleic acid or a variant thereof.

More specifically, the present invention provides a method for the production of transgenic plants having improved growth characteristics, which method comprises:

(iii) introducing into a plant or plant cell an rbp1 nucleic acid or variant thereof; and

(iv) cultivating the plant cell under conditions promoting plant growth and development.

The nucleic acid may be introduced directly into a plant cell or into the plant itself (including introduction into a tissue, organ or any other part of a plant). According to a preferred feature of the present invention, the nucleic acid is preferably introduced into a plant by transformation.

The term "transformation" as referred to herein encompasses the transfer of an exogenous polynucleotide into a host cell, irrespective of the method used for transfer. Plant tissue capable of subsequent clonal propagation, whether by organogenesis or embryogenesis, may be transformed with a genetic construct of the present invention and a whole plant regenerated therefrom. The particular tissue chosen will vary depending on the clonal propagation systems available for, and best suited to, the particular species being transformed. Exemplary tissue targets include leaf disks, pollen, embryos, cotyledons, hypocotyls, megagametophytes, callus tissue, existing meristematic tissue (e.g., apical meristem, axillary buds, and root meristems),

and induced meristem tissue (e.g., cotyledon meristem and hypocotyl meristem). The polynucleotide may be transiently or stably introduced into a host cell and may be maintained non-integrated, for example, as a plasmid. Alternatively, it may be integrated into the host genome. The resulting transformed plant cell may then be used to regenerate a transformed plant in a manner known to persons skilled in the art.

Transformation of plant species is now a fairly routine technique. Advantageously, any of several transformation methods may be used to introduce the gene of interest into a suitable ancestor cell. Transformation methods include the use of liposomes, electroporation, chemicals that increase free DNA uptake, injection of the DNA directly into the plant, particle gun bombardment, transformation using viruses or pollen and microprojection. Methods may be selected from the calcium/polyethylene glycol method for protoplasts (Krens, F.A. *et al.*, 1882, Nature 296, 72-74; Negrutiu I. *et al.*, June 1987, Plant Mol. Biol. 8, 363-373); electroporation of protoplasts (Shillito R.D. *et al.*, 1985 Bio/Technol 3, 1099-1102); microinjection into plant material (Crossway A. *et al.*, 1986, Mol. Gen Genet 202, 179-185); DNA or RNA-coated particle bombardment (Klein T.M. *et al.*, 1987, Nature 327, 70) infection with (non-integrative) viruses and the like. Transgenic rice plants expressing an RNA-binding protein are preferably produced via *Agrobacterium*-mediated transformation using any of the well known methods for rice transformation, such as described in any of the following: published European patent application EP 1198985 A1, Aldemita and Hodges (Planta, 199, 612-617, 1996); Chan *et al.* (Plant Mol. Biol. 22 (3) 491-506, 1993), Hiei *et al.* (Plant J. 6 (2) 271-282, 1994), which disclosures are incorporated by reference herein as if fully set forth. In the case of corn transformation, the preferred method is as described in either Ishida *et al.* (Nat. Biotechnol. 1996 Jun; 14(6): 745-50) or Frame *et al.* (Plant Physiol. 2002 May; 129(1): 13-22), which disclosures are incorporated by reference herein as if fully set forth.

Generally after transformation, plant cells or cell groupings are selected for the presence of one or more markers which are encoded by plant-expressible genes co-transferred with the gene of interest, following which the transformed material is regenerated into a whole plant.

Following DNA transfer and regeneration, putatively transformed plants may be evaluated, for instance using Southern analysis, for the presence of the gene of interest, copy number and/or genomic organisation. Alternatively or additionally, expression levels of the newly introduced DNA may be monitored using Northern and/or Western analysis, both techniques being well known to persons having ordinary skill in the art.

The generated transformed plants may be propagated by a variety of means, such as by clonal propagation or classical breeding techniques. For example, a first generation (or T1) transformed plant may be selfed to give homozygous second generation (or T2) transformants, and the T2 plants further propagated through classical breeding techniques.

The generated transformed organisms may take a variety of forms. For example, they may be chimeras of transformed cells and non-transformed cells; clonal transformants (e.g., all cells transformed to contain the expression cassette); grafts of transformed and untransformed tissues (e.g., in plants, a transformed rootstock grafted to an untransformed scion).

The present invention clearly extends to any plant cell or plant produced by any of the methods described herein, and to all plant parts and propagules thereof. The present invention extends further to encompass the progeny of a primary transformed or transfected cell, tissue, organ or whole plant that has been produced by any of the aforementioned methods, the only requirement being that progeny exhibit the same genotypic and/or phenotypic characteristic(s) as those produced in the parent by the methods according to the invention. The invention also includes host cells containing an isolated RNA-binding protein nucleic acid or variant thereof. Preferred host cells according to the invention are plant cells. The invention also extends to harvestable parts of a plant, such as but not limited to seeds, leaves, fruits, flowers, stem cultures, rhizomes, tubers and bulbs.

The present invention also encompasses the use of RNA-binding protein nucleic acids or variants thereof and to the use of RNA-binding proteins or homologues thereof.

One such use relates to improving the growth characteristics of plants, in particular in improving yield, especially seed yield. The seed yield may include one or more of the following: increased number of (filled) seeds, increased seed weight, increased harvest index, among others.

RNA-binding protein-encoding nucleic acids or variants thereof or RNA-binding proteins or homologues thereof may find use in breeding programmes in which a DNA marker is identified which may be genetically linked to an RNA-binding protein-encoding gene or variant thereof. The RNA-binding protein or variants thereof or RNA-binding proteins or homologues thereof may be used to define a molecular marker. This DNA or protein marker may then be used in breeding programs to select plants having altered growth characteristics. The RNA-binding protein-encoding gene or variant thereof may, for example, be a nucleic acid as represented by any one of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7 and SEQ ID NO: 9.

Allelic variants of an RNA-binding protein-encoding gene/nucleic acid may also find use in marker-assisted breeding programmes. Such breeding programmes sometimes require introduction of allelic variation by mutagenic treatment of the plants, using for example EMS mutagenesis; alternatively, the programme may start with a collection of allelic variants of so called "natural" origin caused unintentionally. Identification of allelic variants then takes place by, for example, PCR. This is followed by a selection step for selection of superior allelic variants of the sequence in question and which give improved growth characteristics in a plant. Selection is typically carried out by monitoring growth performance of plants containing different allelic variants of the sequence in question, for example, different allelic variants of any one of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7 and SEQ ID NO: 9. Growth performance may be monitored in a greenhouse or in the field. Further optional steps include crossing plants, in which the superior allelic variant was identified, with another plant. This could be used, for example, to make a combination of interesting phenotypic features.

RNA-binding protein-encoding nucleic acids or variants thereof may also be used as probes for genetically and physically mapping the genes that they are a part of, and as markers for traits linked to those genes. Such information may be useful in plant breeding in order to develop lines with desired phenotypes. Such use of RNA-binding protein-encoding nucleic acids or variants thereof requires only a nucleic acid sequence of at least 15 nucleotides in length. The RNA-binding protein-encoding nucleic acids or variants thereof may be used as restriction fragment length polymorphism (RFLP) markers. Southern blots (Maniatis) of restriction-digested plant genomic DNA may be probed with the RNA-binding protein-encoding nucleic acids or variants thereof. The resulting banding patterns may then be subjected to genetic analyses using computer programs such as MapMaker (Lander *et al.* (1987) Genomics 1:174-181) in order to construct a genetic map. In addition, the nucleic acids may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted and used to calculate the position of the RNA-binding protein encoding nucleic acid or variant thereof in the genetic map previously obtained using this population (Botstein *et al.* (1980) Am. J. Hum. Genet. 32:314-331).

The production and use of plant gene-derived probes for use in genetic mapping is described in Bematzky and Tanksley (1986) Plant Mol. Biol. Reporter 4:37-41. Numerous publications describe genetic mapping of specific cDNA clones using the methodology outlined above or variations thereof. For example, F2 intercross populations, backcross populations, randomly

mated populations, near isogenic lines, and other sets of individuals may be used for mapping. Such methodologies are well known to those skilled in the art.

The nucleic acid probes may also be used for physical mapping (i.e., placement of sequences on physical maps; see Hoheisel *et al.* In: Non-mammalian Genomic Analysis: A Practical Guide, Academic press 1996, pp. 319-346, and references cited therein).

In another embodiment, the nucleic acid probes may be used in direct fluorescence *in situ* hybridization (FISH) mapping (Trask (1991) Trends Genet. 7:149-154). Although current methods of FISH mapping favor use of large clones (several to several hundred KB; see Laan *et al.* (1995) Genome Res. 5:13-20), improvements in sensitivity may allow performance of FISH mapping using shorter probes.

A variety of nucleic acid amplification-based methods of genetic and physical mapping may be carried out using the nucleic acids. Examples include allele-specific amplification (Kazazian (1989) J. Lab. Clin. Med 11:95-96), polymorphism of PCR-amplified fragments (CAPS; Sheffield *et al.* (1993) Genomics 16:325-332), allele-specific ligation (Landegren *et al.* (1988) Science 241:1077-1080), nucleotide extension reactions (Sokolov (1990) Nucleic Acid Res. 18:3671), Radiation Hybrid Mapping (Walter *et al.* (1997) Nat. Genet. 7:22-28) and Happy Mapping (Dear and Cook (1989) Nucleic Acid Res. 17:6795-6807). For these methods, the sequence of a nucleic acid is used to design and produce primer pairs for use in the amplification reaction or in primer extension reactions. The design of such primers is well known to those skilled in the art. In methods employing PCR-based genetic mapping, it may be necessary to identify DNA sequence differences between the parents of the mapping cross in the region corresponding to the instant nucleic acid sequence. This, however, is generally not necessary for mapping methods.

RNA-binding protein-encoding nucleic acids or variants thereof or RNA-binding proteins or homologues thereof may also find use as growth regulators. Since these molecules have been shown to be useful in improving the growth characteristics of plants, they would also be useful growth regulators, such as herbicides or growth stimulators. The present invention therefore provides a composition comprising an RNA-binding protein-encoding nucleic acid/gene or variant thereof or an RNA-binding protein or homologue thereof, together with a suitable carrier, diluent or excipient, for use as a growth regulator.

The present invention also encompasses the use of rbp1 nucleic acids or variants thereof and to the use of RBP1 polypeptides or homologues thereof.

One such use relates to improving the growth characteristics of plants, in particular in improving yield, especially seed yield. The seed yield may include one or more of the following: increased number of (filled) seeds, increased seed weight, among others.

Rbp1 nucleic acids or variants thereof or RPB1 polypeptides or homologues thereof may find use in breeding programmes in which a DNA marker is identified which may be genetically linked to an rbp1 gene or variant thereof. The rbp1 or variants thereof or RBP1 or homologues thereof may be used to define a molecular marker. This DNA or protein marker may then be used in breeding programs to select plants having altered growth characteristics. The rbp1 gene or variant thereof may, for example, be a nucleic acid as represented by any one of SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34 and SEQ ID NO: 36.

Allelic variants of an rbp1 may also find use in marker-assisted breeding programmes. Such breeding programmes sometimes require introduction of allelic variation by mutagenic treatment of the plants, using for example EMS mutagenesis; alternatively, the programme may start with a collection of allelic variants of so called "natural" origin caused unintentionally. Identification of allelic variants then takes place by, for example, PCR. This is followed by a selection step for selection of superior allelic variants of the sequence in question and which give rise improved growth characteristics in a plant. Selection is typically carried out by monitoring growth performance of plants containing different allelic variants of the sequence in question, for example, different allelic variants of any one of SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34 and SEQ ID NO: 36. Growth performance may be monitored in a greenhouse or in the field. Further optional steps include crossing plants, in which the superior allelic variant was identified, with another plant. This could be used, for example, to make a combination of interesting phenotypic features.

An rbp1 nucleic acid or variant thereof may also be used as probes for genetically and physically mapping the genes that they are a part of, and as markers for traits linked to those genes. Such information may be useful in plant breeding in order to develop lines with desired phenotypes. Such use of rbp1 nucleic acids or variants thereof requires only a nucleic acid sequence of at least 15 nucleotides in length. The rbp1 nucleic acids or variants thereof may be used as restriction fragment length polymorphism (RFLP) markers. Southern blots (Maniatis) of restriction-digested plant genomic DNA may be probed with the rbp1 nucleic

acids or variants thereof. The resulting banding patterns may then be subjected to genetic analyses using computer programs such as MapMaker (Lander et al. (1987) Genomics 1:174-181) in order to construct a genetic map. In addition, the nucleic acids may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted and used to calculate the position of the rbp1 nucleic acid or variant thereof in the genetic map previously obtained using this population (Botstein *et al.* (1980) Am. J. Hum. Genet. 32:314-331).

The production and use of plant gene-derived probes for use in genetic mapping is described in Bematzky and Tanksley (1986) Plant Mol. Biol. Reporter 4:37-41. Numerous publications describe genetic mapping of specific cDNA clones using the methodology outlined above or variations thereof. For example, F2 intercross populations, backcross populations, randomly mated populations, near isogenic lines, and other sets of individuals may be used for mapping. Such methodologies are well known to those skilled in the art.

The nucleic acid probes may also be used for physical mapping (i.e., placement of sequences on physical maps; see Hoheisel *et al.* In: Nonmammalian Genomic Analysis: A Practical Guide, Academic press 1996, pp. 319-346, and references cited therein).

In another embodiment, the nucleic acid probes may be used in direct fluorescence *in situ* hybridization (FISH) mapping (Trask (1991) Trends Genet. 7:149-154). Although current methods of FISH mapping favor use of large clones (several to several hundred KB; see Laan *et al.* (1995) Genome Res. 5:13-20), improvements in sensitivity may allow performance of FISH mapping using shorter probes.

A variety of nucleic acid amplification-based methods of genetic and physical mapping may be carried out using the nucleic acids. Examples include allele-specific amplification (Kazazian (1989) J. Lab. Clin. Med 11:95-96), polymorphism of PCR-amplified fragments (CAPS; Sheffield et al. (1993) Genomics 16:325-332), allele-specific ligation (Landegren *et al.* (1988) Science 241:1077-1080), nucleotide extension reactions (Sokolov (1990) Nucleic Acid Res. 18:3671), Radiation Hybrid Mapping (Walter et al. (1997) Nat. Genet. 7:22-28) and Happy Mapping (Dear and Cook (1989) Nucleic Acid Res. 17:6795-6807). For these methods, the sequence of a nucleic acid is used to design and produce primer pairs for use in the amplification reaction or in primer extension reactions. The design of such primers is well known to those skilled in the art. In methods employing PCR-based genetic mapping, it may be necessary to identify DNA sequence differences between the parents of the mapping cross

in the region corresponding to the instant nucleic acid sequence. This, however, is generally not necessary for mapping methods.

rbp1 nucleic acids or variants thereof or RBP1 polypeptides or homologues thereof may also find use as growth regulators. Since these molecules have been shown to be useful in improving the growth characteristics of plants, they would also be useful growth regulators, such as herbicides or growth stimulators. The present invention therefore provides a composition comprising an rbp1 or variant thereof or an RBP1 polypeptide or homologue thereof, together with a suitable carrier, diluent or excipient, for use as a growth regulator.

The methods according to the present invention result in plants having improved growth characteristics, as described hereinabove. These advantageous growth characteristics may also be combined with other economically advantageous traits, such as further yield-enhancing traits, tolerance to various stresses, traits modifying various architectural features and/or biochemical and/or physiological features.

Description of figures

The present invention will now be described with reference to the following figures in which:

Fig. 1 shows a CLUSTAL multiple alignment of plant RNA-binding proteins. Motifs I and II are boxed (M2 is absent from BAC83046) and RRM domains are underlined.

Fig. 2 shows a binary vector for expression in *Oryza sativa* of a tobacco RNA-binding protein under the control of a prolamin promoter.

Fig. 3 shows a multiple alignment of plant RBP1 polypeptides. Genebank protein or their encoding nucleic acids are indicated. At denotes *Arabidopsis thaliana* and Os denotes *oryza sativa*.

Fig. 4 shows a binary vector for expression in *Oryza sativa* of an *Arabidopsis thaliana* RBP1 (internal reference CDS0078) under the control of a beta expansin promoter (internal reference PRO0061).

Fig. 5 shows photographs of GUS expression driven by a beta expansin promoter. The photograph of the "C plant" is of a rice plant GUS stained when it had reached a size of about 5 cm. The photograph of the "B plant" is of a rice plant GUS stained when it had reached a

size of about 10 cm. Promoters with comparable expression profiles may also be useful in the methods of the invention.

Fig. 6 details examples of sequences useful in performing the methods according to the present invention. From SEQ ID NO: 14 onwards, the At number given refers to the MIPs Accession number (http://mips.gsf.de/); other identifiers refer to Genbank accession numbers. Capital letters represent the coding sequence and small letters refer to non-translated regions, including 5' leader sequences, 3' untranslated regions and introns. Chromosomic location of the gene is indicated by the contig number and coordinates of the ORF in the contig.

Examples

The present invention will now be described with reference to the following examples, which are by way of illustration alone.

DNA manipulation: unless otherwise stated, recombinant DNA techniques are performed according to standard protocols described in (Sambrook (2001) Molecular Cloning: a laboratory manual, 3rd Edition Cold Spring Harbor Laboratory Press, CSH, New York) or in Volumes 1 and 2 of Ausubel *et al.* (1994), Current Protocols in Molecular Biology, Current Protocols. Standard materials and methods for plant molecular work are described in Plant Molecular Biology Labfase (1993) by R.D.D. Croy, published by BIOS Scientific Publications Ltd (UK) and Blackwell Scientific Publications (UK).

Example 1: Gene Cloning - tobacco RNA-binding protein-encoding gene

A gene encoding an RNA-binding protein was first identified as an expressed sequence tag from Tobacco BY2 cells and was isolated as a partial sequence in a CDNA-AFLP experiment performed with cDNA made from a synchronized tobacco BY2 cell culture (*Nicotiniana tabacum* L. cv. Bright Yellow-2). Based on this cDNA-AFLP experiment, BY2 tags that were cell cycle modulated were identified and selected for further cloning. The expressed sequence tags were used to screen a Tobacco cDNA library and to isolate the full length cDNA.

Synchronization of BY2 cells

Tobacco BY2 (*Nicotiana tabacum* L. cv. Bright Yellow-2) cultured cell suspension was synchronized by blocking cells in early S-phase with aphidicolin as follows. A cultured cell suspension of *Nicotiana tabacum* L. cv. Bright Yellow 2 was maintained as described (Nagata *et al.* Int. Rev. Cytol. 132, 1-30, 1992). For synchronization, a 7-day-old stationary culture was diluted 10-fold in fresh medium supplemented with aphidicolin (Sigma-Aldrich, St. Louis, MO;

5 mg/l), a DNA-polymerase α inhibiting drug. After 24 h, the cells were released from the block by several washings with fresh medium and they resumed their cell cycle progression.

RNA extraction and cDNA synthesis

Total RNA was prepared using LiCl precipitation (Sambrook *et al.*, 2001) and poly(A^+) RNA was extracted from 500 μg of total RNA using Oligotex columns (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Starting from 1 μg of poly(A^+) RNA, first-strand cDNA was synthesized by reverse transcription with a biotinylated oligo-dT_{25} primer (Genset, Paris, France) and Superscript II (Life Technologies, Gaithersburg, MD). Second-strand synthesis was done by strand displacement with *Escherichia coli* ligase (Life Technologies), DNA polymerase I (USB, Cleveland, OH) and RNAse-H (USB).

cDNA-AFLP analysis

Five hundred ng of double-stranded cDNA was used for AFLP analysis as described (Vos *et al.*, Nucleic Acids Res. 23 (21) 4407-4414, 1995; Bachem *et al.*, Plant J. 9 (5) 745-53, 1996) with modifications. The restriction enzymes used were *Bst*YI and *Mse*I (Biolabs) and the digestion was done in two separate steps. After the first restriction digest with one of the enzymes, the 3' end fragments were collected on Dyna beads (Dynal, Oslo, Norway) by means of their biotinylated tail, while the other fragments were washed away. After digestion with the second enzyme, the released restriction fragments were collected and used as templates in the subsequent AFLP steps. For preamplifications, a *Mse*I primer without selective nucleotides was combined with a *Bst*YI primer containing either a T or a C as 3' most nucleotide. PCR conditions were as described (Vos *et al.*, 1995). The obtained amplification mixtures were diluted 600-fold and 5 μl was used for selective amplifications using a P^{33}-labeled *Bst*YI primer and the Amplitaq-Gold polymerase (Roche Diagnostics, Brussels, Belgium). Amplification products were separated on 5% polyacrylamide gels using the Sequigel system (Biorad). Dried gels were exposed to Kodak Biomax films as well as scanned in a phosphoImager (Amersham Pharmacia Biotech, Little Chalfont, UK).

Characterization of AFLP fragments

Bands corresponding to differentially expressed transcripts, among which was the transcript corresponding to SEQ ID NO 1, were isolated from the gel and eluted DNA was reamplified under the same conditions as for selective amplification. Sequence information was obtained either by direct sequencing of the reamplified polymerase chain reaction product with the selective *Bst*YI primer or after cloning the fragments in pGEM-T easy (Promega, Madison, WI) or by sequencing individual clones. The obtained sequences were compared against nucleotide and protein sequences present in the publicly available databases by BLAST

sequence alignments (Altschul *et al.*, Nucleic Acids Res. 25 (17) 3389-3402 1997). When available, tag sequences were replaced with longer EST or isolated cDNA sequences to increase the chance of finding significant homology. The physical cDNA clone corresponding to SEQ ID NO 1 was subsequently amplified from a commercial Tobacco cDNA library as follows.

Gene Cloning

A c-DNA library with average inserts of 1,400 bp was made with poly(A^+) isolated from actively dividing, non-synchronized BY2 tobacco cells. These library-inserts were cloned in the vector pCMVSPORT6.0, comprising a attB gateway cassette (Life Technologies). From this library 46,000 clones were selected, arrayed in 384-well microtiter plates, and subsequently spotted in duplicate on nylon filters. The arrayed clones were screened by using pools of several hundreds of radioactively labeled tags as probes (among which was the BY2-tag corresponding to the sequence of SEQ IDNO 1). Positive clones were isolated (among which the clone reacting with the BY2-tag corresponding to the sequence of SEQ ID NO 1), sequenced, and aligned with the tag sequence. In cases where hybridisation with the tag failed, the full-length cDNA corresponding to the tag was selected by PCR amplification as follows. Tag-specific primers were designed using primer3 program (http://www-genome.wi.mit.edu/genome_software/other/primer3.html) and used in combination with the common vector primer to amplify partial cDNA inserts. Pools of DNA, from 50,000, 100,000, 150,000, and 300,000 cDNA clones were used as templates in PCR amplifications. Amplification products were isolated from agarose gels, cloned, sequenced and aligned with tags.

Subsequently, the full-length cDNA corresponding to SEQ ID NO 1 was cloned from the pCMVsport6.0 library vector into a suitable plant expression vector via an LR Gateway reaction.

LR gateway reaction to clone CDS0701 into a plant expression vector

The pCMV Sport 6.0 p2461 was subsequently used in an LR reaction with a Gateway destination vector suitable for rice transformation. This vector contains as functional elements within the T-DNA borders a plant selectable marker and a Gateway cassette intended for LR *in vivo* recombination with the sequence of interest already cloned in the donor vector. Upstream of this Gateway cassette is the rice prolamin promoter for seed specific expression of the gene.

After the recombination step, the resulting expression vector (see Fig. 2) was transformed into *Agrobacterium* strain LBA4404 and subsequently into rice plants.

Example 2: Rice Transformation

Mature dry seeds of the rice japonica cultivar Nipponbare (NB) were dehusked. Sterilization was carried out by incubating for one minute in 70% ethanol, followed by 30 minutes in 0.2%HgCl2, followed by a 6X15 minute wash with sterile distilled water. The sterile seeds were then germinated on a medium containing 2,4-D (callus induction medium). After incubation in the dark for four weeks, embryogenic, scutellum-derived calli were excised and propagated on the same medium. After two weeks, the calli were multiplied or propagated by subculture on the same medium for another 2 weeks. Embryogenic callus pieces were sub-cultured on fresh medium 3 days before co-cultivation (to boost cell division activity). *Agrobacterium* strain LBA4404 harbouring binary T-DNA vectors were used for co-cultivation. *Agrobacterium* was inoculated on AB medium with the appropriate antibiotics and cultured for 3 days at 28°C. The bacteria were then collected and suspended in liquid co-cultivation medium to a density (OD600) of about 1. The suspension was then transferred to a petri dish and the calli immersed in the suspension for 15 minutes. The callus tissues were then blotted dry on a filter paper and transferred to solidified, co-cultivation medium and incubated for 3 days in the dark at 25°C. Co-cultivated calli were grown on 2,4-D-containing medium for 4 weeks in the dark at 28°C in the presence of a suitable concentration of the selective agent. During this period, rapidly growing resistant callus islands developed. After transfer of this material to a regeneration medium and incubation in the light, the embryogenic potential was released and shoots developed in the next four to five weeks. Shoots were excised from the calli and incubated for 2 to 3 weeks on an auxin-containing medium from which they were transferred to soil. Hardened shoots were grown under high humidity and short days in a greenhouse. Seeds were then harvested three to five months after transplanting. The method yielded single locus transformants at a rate of over 50 (Aldemita and Hodges, Planta, 199 612-617, 1996; Chan *et al.*, Plant Mol. Biol. 22 (3) 491-506, 1993, Hiei *et al.*, Plant J., 6 (2) 271-282, 1994).

Example 3: Evaluation and Results

Approximately 15 to 20 independent T0 rice transformants were generated. The primary transformants were transferred from tissue culture chambers to a greenhouse for growing and harvest of T1 seed. 5 events, of which the T1 progeny segregated 3:1 for presence/absence of the transgene, were retained. For each of these events, approximately 10 T1 seedlings containing the transgene (hetero- and homo-zygotes), and in the same number, approximately 10 T1 seedlings lacking the transgene (nullizygotes), were selected by monitoring visual

marker expression. 4 T1 events were further evaluated in the T2 generation following the same evaluation procedure as for the T1 generation but with more individuals per event.

Statistical analysis: F-test

A two factor ANOVA (analysis of variants) was used as a statistical model for the overall evaluation of plant phenotypic characteristics. An F-test was carried out on all the parameters measured of all the plants of all the events transformed with the gene of the present invention. The F-test was carried out to check for an effect of the gene over all the transformation events and to verify for an overall effect of the gene, also known as a global gene effect. The threshold for significance for a true global gene effect was set at a 5% probability level for the F-test. A significant F-test value points to a gene effect, meaning that it is not only the presence or position of the gene that is causing the differences in phenotype.

3.1 Seed-related parameter measurements

The mature primary panicles were harvested, bagged, barcode-labelled and then dried for three days in the oven at 37°C. The panicles were then threshed and all the seeds were collected and counted. The filled husks were separated from the empty ones using an air-blowing device. The empty husks were discarded and the remaining fraction was counted again. The filled husks were weighed on an analytical balance. The total seed yield was measured by weighing all filled husks harvested from a plant. The harvest index in the present invention is defined as a ratio of total seed yield and the aboveground area (mm^2) multiplied by a factor 10^6.

The Table of results below show the p values from the F test for T1 and T2 evaluations. The percentage difference between the transgenics and the corresponding nullizygotes is also shown. For example, for total seed weight in the T1 generation, 3 out of 4 lines were positive for total seed weight (i.e., showed an increase in total seed weight (of greater than 32%) compared to the seed weight of corresponding nullizygote plants). 2 out of 4 of these lines showed a significant increase in total seed weight with a p value from the F test of 0.061.

Table 5: Results of the T1 generation

T1	Number of lines showing an increase	Difference	Number of lines showing a significant increase	p value of F test
Total weight seeds	3 out 4	> 32%	2 out 4	< 0.061
Harvest index	2 out 4	> 32%	2 out 4	< 0.09

Table 6: results of the T2 generation

T2	Number of lines showing an increase	Difference	Number of lines showing a significant increase	p value of F test
Total weight seeds	1 out 4	> 30%	1 out 4	< 0.064
Harvest index	1 out 4	> 40%	1 out 4	< 0.001

Example 4: Gene Cloning AtRBP1

The *Arabidopsis AtRBP1 (CDS0078)* was amplified by PCR using as template an *Arabidopsis thaliana* seedling cDNA library (Invitrogen, Paisley, UK). After reverse transcription of RNA extracted from seedlings, the cDNAs were cloned into pCMV Sport 6.0. Average insert size of the bank was 1.5 kb, and original number of clones was of $1.59x10^7$ cfu. Original titer was determined to be $9.6x10^5$ cfu/ml, after first amplification of $6x10^{11}$ cfu/ml. After plasmid extraction, 200 ng of template was used in a 50 µl PCR mix. Primers prm00405 (sense 5' ggggacaagtttgtacaaaaaagcaggcttcacaatggattatgatcggtacaagttat 3') and prm00406 (reverse, complementary: 5' ggggaccactttgtacaagaaagctgggtttaaaagagtccaaagaatttcact 3'), which include the AttB sites for Gateway recombination, were used for PCR amplification. PCR was performed using Hifi Taq DNA polymerase in standard conditions. A PCR fragment of 1209 bp was amplified and purified also using standard methods. The first step of the Gateway procedure, the BP reaction, was then performed, during which the PCR fragment recombines *in vivo* with the pDONR201 plasmid to produce, according to the Gateway terminology, an "entry clone", p00733. Plasmid pDONR201 was purchased from Invitrogen, as part of the Gateway® technology.

Example 5: Vector Construction AtRBP1

The entry clone p00733 was subsequently used in an LR reaction with p03069, a destination vector used for *Oryza sativa* transformation. This vector contained as functional elements within the T-DNA borders: a plant selectable marker; a visual marker expression cassette; and a Gateway cassette intended for LR *in vivo* recombination with the sequence of interest already cloned in the entry clone. A Beta-Expansin promoter for expression in shoots was located upstream of this Gateway cassette.

After the LR recombination step, the resulting expression vector p04280 (Figure 2) was transformed into the *Agrobacterium* strain LBA4404 and subsequently to *Oryza sativa* plants. Transformed rice plants were allowed to grow and were then examined for the parameters described in Example 6.

Example 6: Evaluation and Results AtRBP1

Approximately 15 to 20 independent T0 rice transformants were generated. The primary transformants were transferred from tissue culture chambers to a greenhouse for growing and harvest of T1 seed. 5 events, of which the T1 progeny segregated 3:1 for presence/absence of the transgene, were retained. For each of these events, approximately 10 T1 seedlings containing the transgene (hetero- and homo-zygotes), and in the same number, approximately 10 T1 seedlings lacking the transgene (nullizygotes), were selected by monitoring visual marker expression. 4 T1 events were further evaluated in the T2 generation following the same evaluation procedure as for the T1 generation but with more individuals per event. One line that was neutral in the first round was not taken along. In the T2 evaluation, 15T2 seedlings containing the transgene are compared to the same number of plants lacking the transgene (nullizygotes).

Statistical analysis: F-test

A two factor ANOVA (analysis of variants) was used as a statistical model for the overall evaluation of plant phenotypic characteristics. An F-test was carried out on all the parameters measured of all the plants of all the events transformed with the gene of the present invention. The F-test was carried out to check for an effect of the gene over all the transformation events and to verify for an overall effect of the gene, also known as a global gene effect. The threshold for significance for a true global gene effect was set at a 5% probability level for the F-test. A significant F-test value points to a gene effect, meaning that it is not only the presence or position of the gene that is causing the differences in phenotype.

6.1 Seed-related parameter measurements

The mature primary panicles were harvested, bagged, barcode-labelled and then dried for three days in the oven at 37°C. The panicles were then threshed and all the seeds were collected and counted. The filled husks were separated from the empty ones using an air-blowing device. The empty husks were discarded and the remaining fraction was counted again. The filled husks were weighed on an analytical balance. This procedure resulted in the set of seed-related parameters described below.

The Table of results below show the p values from the F test for the T1 evaluations, the T2 evaluations and the combined p values from the F tests for the T1 and T2 evaluations. A combined analysis may be considered when two experiments have been carried out on the same events. This may be useful to check for consistency of the effects over the two experiments and to increase confidence in the conclusion. The method used is a mixed-model approach that takes into account the multilevel structure of the data (i.e. experiment - event -

segregants). P-values are obtained by comparing likelihood ratio test to chi square distributions. Each of the tables also gives the % difference between the transgenics and the corresponding nullizygotes for each generation.

6.1.1 Aboveground Area

Plant aboveground area was determined by counting the total number of pixels from aboveground plant parts discriminated from the background. This value was averaged for the pictures taken on the same time point from the different angles and was converted to a physical surface value expressed in square mm by calibration. Experiments show that the aboveground plant area measured this way correlates with the biomass of plant parts above ground. The results of the T1 and T2 evaluation are shown in Table 7 below. As shown in the table below, the p value from the F test for the T2 evaluation (p value of 0.0011) and the combined data (with a p value of 0.0287) were significant indicating that the presence of the construct in the plants has a significant positive effect on aboveground area of transgenic plants.

Table 7: Aboveground Area

Aboveground area		
	% Difference	P value
T1 Overall	8	0.1779
T2 Overall	15	0.0011
Combined		0.0012

6.1.2 Total seed yield per plant

The total seed yield was measured by weighing all filled husks harvested from a plant. As shown in Table 8 below, the p value from the F test for the T1 and T2 evaluation combined was significant (with a p value of 0.0287) indicating that the presence of the construct in the plants has a significant effect on the total seed weight of transgenic plants.

Table 8

Total Seed Weight		
	% Difference	P value
T1	12	0.3397
T2	16	0.1356
Combined		0.0287

6.1.3 Total Number of Seeds

As shown in Table 9 below, the p value from the F test for the T1 and T2 evaluation combined (and T2 individually) was significant (with a p value of 0.0006) indicating that the presence of the construct in the plants has a significant effect on the total number of seeds of transgenic plants.

Table 9

Total Number of seeds		
	% Difference	P value
T1	6	0.4044
T2	23	0.0003
Combined		0.0006

Example 7: GUS expression driven by beta expansin promoter

The beta-expansin promoter was cloned into the pDONR201 entry plasmid of the GatewayTM system (Life Technologies) using the "BP recombination reaction". The identity and base pair composition of the cloned insert was confirmed by sequencing and additionally, the resulting plasmid was tested via restriction digests.

In order to clone the promoter in front of a reporter gene, each entry clone was subsequently used in an "LR recombination reaction" (Gateway TM) with a destination vector. This destination vector was designed to operably link the promoter to the *Escherichia coli* beta-glucuronidase (GUS) gene via the substitution of the Gateway recombination cassette in front of the GUS gene. The resulting reporter vectors, comprising the promoter operably linked to GUS were subsequently transformed into *Agrobacterium* strain LBA4044 and subsequently into rice plants using standard transformation techniques.

Transgenic rice plants were generated from transformed cells. Plant growth was performed under normal conditions.

The plants or plant parts to be tested were covered with 90% ice-cold acetone and incubated for 30 min at 4 °C. After 3 washes of 5 min with Tris buffer [15,76 g Trizma HCl (Sigma T3253) + 2,922 g NaCl in 1 litre bi-distilled water, adjusted to pH 7,0 with NaOH], the material was covered by a Tris/ferricyanate/X-Gluc solution [9,8 ml Tris buffer + 0,2 ml ferricyanate stock (0,33 g Potassium ferricyanate (Sigma P3667) in 10 ml Tris buffer)+ 0,2 ml X-Gluc stock (26,1 mg X-Gluc (Europa Bioproducts ML 113A) in 500 µl DMSO)]. Vacuum infiltration was applied

for 15 to 30 minutes. The plants or plant parts were incubated for up to 16 hours at 37 °C until development of blue colour was visible. The samples were washed 3 times for 5 minutes with Tris buffer. Chlorophyll was extracted in ethanol series of 50%, 70% and 90% (each for 30 minutes).

5

Claims

1. Method for improving plant growth characteristics, comprising increasing activity in a plant of an RNA-binding protein or a homologue thereof, wherein said RNA-binding protein or homologue thereof is either:
   (i) a polypeptide having RNA-binding activity and comprising either 2 or 3 RNA recognition motifs (RRMs) and a motif having at least 75% sequence identity to motif I: PYEAAVVALPVVVKERLVRILRLGIATRYD and/or a motif having at least 50% sequence identity to motif II: RFDPFTGEPYKFDP; or
   (ii) an RBP1 polypeptide or homologue thereof having (a) RNA-binding activity; (b) two RRM domains, (c) the following two motifs: (i) KIFVGGL; and (ii) RPRGFGF, allowing for up to three amino acid substitutions and any conservative change in the motifs; and (d) having, in increasing order of preference, at least 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% sequence identity to the amino acid represented by SEQ ID NO: 15; and
   (iii) optionally selecting for plants having improved growth characteristics.

2. Method according to claim 1, wherein said increased activity is effected by introducing a genetic modification preferably in the locus of a gene encoding an RNA-binding protein or a homologue thereof.

3. Method according to claim 2, wherein said genetic modification is effected by one of site-directed mutagenesis, homologous recombination, TILLING or T-DNA activation.

4. Method for improving plant growth characteristics, comprising introducing and expressing in a plant an RNA-binding protein-encoding nucleic acid or a functional variant thereof, wherein said encoded RNA-binding protein or homologue thereof is either:
   (i) a polypeptide having RNA-binding activity and comprising either 2 or 3 RNA recognition motifs (RRMs) and a motif having at least 75% sequence identity to motif I: PYEAAVVALPVVVKERLVRILRLGIATRYD and/or a motif having at least 50% sequence identity to motif II: RFDPFTGEPYKFDP; or
   (ii) an RBP1 polypeptide or homologue thereof having (a) RNA-binding activity; (b) two RRM domains, (c) the following two motifs: (i) KIFVGGL; and (ii) RPRGFGF, allowing for up to three amino acid substitutions and any conservative change in the motifs; and (d) having, in increasing order of preference, at least 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%,

65%, 70%, 75%, 80%, 85%, 90%, 95% sequence identity to the amino acid represented by SEQ ID NO: 15.

5. Method according to claim 4, wherein said variant is a portion of an RNA-binding protein-encoding nucleic acid or a sequence capable of hybridising to an RNA-binding protein-encoding nucleic acid.

6. Method according to claim 4 or 5, wherein said RNA-binding protein-encoding nucleic acid or functional variant thereof is overexpressed in a plant.

7. Method according to any one of claims 4 to 6, wherein said RNA-binding protein-encoding nucleic acid or functional variant thereof of claim 4(i) is of plant origin, preferably from a dicotyledonous plant, further preferably from the family *Nicotianae*, more preferably the nucleic acid is from tobacco.

8. Method according to any one of claims 4 to 6, wherein said RNA-binding protein-encoding nucleic acid or functional variant thereof of claim 4(ii) is of plant origin, preferably from a dicotyledonous plant, further preferably from the family *Brassicaceae*, more preferably the nucleic acid is from *Arabidopsis thaliana*.

9. Method according to any one of claims 4 to 7, wherein said RNA-binding protein-encoding nucleic acid or functional variant thereof is operably linked to a seed-preferred promoter.

10. Method according to claim 9, wherein said promoter is a prolamin promoter, such a prolamin promoter from rice.

11. Method according to any one of claims 4 to 6 and 8, wherein said rbp1 nucleic acid or functional variant thereof is operably linked to a promoter capable of preferentially expressing said nucleic acid in shoots.

12. Method according to claim 11, wherein said promoter has a comparable expression profile to a beta-expansin promoter.

13. Method according to any one of claims 1 to 12, wherein said improved plant growth characteristic is increased yield relative to corresponding wild type plants.

14. Method according to any one of claims 1 to 13, wherein said improved plant growth characteristic is increased plant biomass.

15. Method according to claim 13, wherein said increased yield is increased seed yield.

16. Method according to claim 15, wherein said increased seed yield is selected from any one or more of (i) increased seed biomass; (ii) increased number of (filled) seeds; (iii) increased seed size; (iv) increased seed volume; (v) increased harvest index; and (vi) increased thousand kernel weight (TKW).

17. Method according to any of one of claims 1 to 16, wherein said improved plant growth characteristic is increased growth rate.

18. Plants obtainable by a method according to any of claims 1 to 17.

19. Construct comprising:

(i) An RNA-binding protein-encoding nucleic acid or variant thereof, which RNA-binding protein has RNA-binding activity and comprises either 2 or 3 RNA recognition motifs (RRMs) and a motif having at least 75% sequence identity to motif I: PYEAAVVALPVVVKERLVRILRLGIATRYD and/or a motif having at least 50% sequence identity to motif II: RFDPFTGEPYKFDP;

(ii) one or more control sequence capable of driving expression of the nucleic acid sequence of (i); and optionally

(iii) a transcription termination sequence.

20. Construct comprising:

(i) An rbp1-encoding nucleic acid or variant thereof, wherein said encoded RBP1 has: (a) RNA-binding activity; (b) two RRM domains, (c) the following two motifs: (i) KIFVGGL; and (ii) RPRGFGF, allowing for up to three amino acid substitutions and any conservative change in the motifs; and (d) having, in increasing order of preference, at least 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% sequence identity to the amino acid represented by SEQ ID NO: 15;

(ii) one or more control sequence capable of driving expression of the nucleic acid sequence of (i); and optionally

(iii) a transcription termination sequence.

21. Construct according to claim 19, wherein said control sequence is a promoter capable of driving expression in seed tissue.

22. Construct according to claim 21, wherein said promoter is a prolamin promoter, such as a prolamin promoter from rice.

23. Construct according to claim 20, wherein said promoter is capable of driving expression in shoots.

24. Construct according to claim 23, wherein said promoter has a comparable expression profile to a beta-expansin promoter.

25. Plant transformed with a construct according to any one of claims 19 to 24.

26. Method for the production of a transgenic plant having improved growth characteristics, which method comprises:

(i) introducing into a plant an RNA-binding protein-encoding nucleic acid or variant thereof, which encoded RNA-binding protein has RNA-binding activity and comprises either 2 or 3 RNA recognition motifs (RRMs) and a motif having at least 75% sequence identity to motif I: PYEAAVVALPVVVKERLVRILRLGIATRYD and/or a motif having at least 50% sequence identity to motif II: RFDPFTGEPYKFDP; and

(ii) cultivating the plant cell under conditions promoting plant growth and development.

27. Method for the production of a transgenic plant having modified growth characteristics, which method comprises:

(i) introducing into a plant an rbp1-encoding nucleic acid or variant thereof, wherein said encoded RBP1 has: (a) RNA-binding activity; (b) two RRM domains, (c) the following two motifs: (i) KIFVGGL; and (ii) RPRGFGF, allowing for up to three amino acid substitutions and any conservative change in the motifs; and (d) having, in increasing order of preference, at least 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% sequence identity to the amino acid represented by SEQ ID NO: 15; and

(ii) cultivating the plant cell under conditions promoting plant growth and development.

28. Transgenic plant having improved growth characteristics resulting from an RNA-binding protein-encoding nucleic acid or variant thereof introduced into said plant.

29. Transgenic plant having improved growth characteristics resulting from an rbp1 nucleic acid or variant thereof introduced into said plant.

30. Transgenic plant according to claim 18, 25 or 28 and 29, wherein said plant is a monocotyledonous plant, such as sugarcane or wherein the plant is a cereal, such as rice, maize, wheat, barley, millet, rye oats or sorghum.

31. Harvestable parts of a plant according to any one of claims 18, 25 or 28 to 30.

32. Harvestable parts according to claim 31, wherein said harvestable parts are seeds.

33. Use of an RNA-binding protein-encoding nucleic acid/gene or variant thereof or use of an RNA-binding protein or homologue thereof in modifying the growth characteristics of plants, in particular in improving yield, especially seed yield.

34. Use of an rbp1 or variant thereof or use of an RBP1 polypeptide or homologue thereof in modifying the growth characteristics of plants, in particular in improving yield, especially seed yield.

35. Use according to claim 33 or 34, wherein said seed yield includes one or more of the following: increased number of (filled) seeds, increased seed weight, increased harvest index and increased TKW.

36. Use of an RNA-binding protein or variant thereof or use of an rbp1 or variant thereof as a molecular marker.

```
newriceCDS701homologue      MVRARDSIREILPVFSIQSALGTADSAPAIRPVAAASDLVRISSEKSRLD 50
rice                        --------------------------------------------------
maize                       --------------------------------------------------
CDS701Proteinprediction     -----------------------------------MSRLIEHHLANNK 13
BAC83046.1                  ----------------------------------MEPTRRCVPGHLATAA 16

newriceCDS701homologue      LPVPLFFFVARG---GFQEKRGAASHGDYDEQGYWMGFFSLIPRASLALR 97
rice                        --------------------------------------------------
maize                       --------------------------------------------------
CDS701Proteinprediction     QDMKGTEVFVGG---LARTTTESKIHEVFSSCGEIVEIRLIKDQTGVPKG 60
BAC83046.1                  AAAAASPFSPPPSLPLPSALMPPKKRRLFTPAPRHAATPPPPPPPPTPAV 66

newriceCDS701homologue      GRRVKGAEVFVGGLPRSVTERALREVGVLPRSQQVFSPCGEIVDLRIMKD 147
rice                        --------------------------------------------------
maize                       --------------------------------------------------
CDS701Proteinprediction     FCFVRFATKYAADKALKEKSGYVLDGKKLGVRPSVEQDTLFLGNLNKGWG 110
BAC83046.1                  EPTLPIPPASTPPTPPQPSASTEPSTAPPPAVDDAAARSSSSSSPASAAA 116

newriceCDS701homologue      QNGISKVLCQLQGKRLAVDLSLDQDTLFFGNLCKGSDWGIEEFEELIRKV 197
rice                        --------------------------------------------------
maize                       --------------------------------------------------
CDS701Proteinprediction     AEEFESIVRQVFPDVVSVDLALLGDVQPG----QKQRNRGFAFVKFPSHA 156
BAC83046.1                  ARKVRKVVKKVIVKKVVPKGTFAAR-----------KAAAAAVAAAAAVS 155

newriceCDS701homologue      RPVVDLAMARNHDSSVGKRRLNRGFAFVRFSSHAVSQVKTAFVGNLPANV 247
rice                        ---------------------------------MAKVKTAFVGNLPANV 16
maize                       GSRTDFMLG-----DILHPAINWADKESHLDPDEMAKMKSAFIGNLPEDV 45
CDS701Proteinprediction     AAARAFRVGSQSDFLIDGKLHPSVQWAEEPDPNELAQIKAAFVRNVPPGA 206
BAC83046.1                  GAAASSEAGGEAPTDEPASDQDGGVGNEQKLDESKPATDCNAVAVVEESV 205
                                                          .  .   :  :  ..

newriceCDS701homologue      TEEYLRKLFEHCGEVCYAVVRVAVSRKGQYPVGFVHFASRTWK------E 291
rice                        TEEYLRKLFEHCGEV----VRVAVSRKGQYPVGFVHFASRT--------E 54
maize                       NEEYLRKLFGQFGEV----VRVAISRKGQCPVAFVHFAKRS--------E 83
CDS701Proteinprediction     DEDYLKKLFQPFGNV----ERIALSRKGSSTIGFVYFDKRS--------D 244
BAC83046.1                  CKEEEEVALVVGKGVEEEEAGMSERRKRMTMEVFVGGLHRDAKEDDVRAV 255
                             ::  .  :     *      ::  **      **    *

newriceCDS701homologue      LDNAIKEMDGETVRGPDRGATFRIQVSVARPVVENDKKRIREEVKTRRSN 341
rice                        LDNAIKEMDGETVRGPDRGATFRIQVSVARPVVENDKKRIREEVKTRRSN 104
maize                       LENAIEEMDGKTVRGPGRGPSFKIQVSVARPTADNDKKRSREEVRTRRSN 133
CDS701Proteinprediction     LDNAIMALNEKTVQGPMGGPSCKLQVEVARPMDKN-RKRGREDPNMSSTI 293
BAC83046.1                  FAKAGEITEVRMIMNPLAGKNKGYCFVRYRHAAQAKKAIAEFGNVKICGK 305
                            : :*    : . : .*  * .    .   *   .  :   .

newriceCDS701homologue      VSTDKPDHSYGRRGHDSYDRQAKAPRLYNEVLHTNDKVDMITLFCQFSFL 391
rice                        VSTDKPDHSYGRRGHDSYDRQAKAPRLYN---------------------- 133
maize                       ASGDRRDYSHGRYGHDSLDRQVKAPRLSN--------------------Y 163
CDS701Proteinprediction     ESHSKLLK------DDPDVEMIRAPKSTAQ--------------------L 318
BAC83046.1                  LCRAAVPVGNDRIFLGNINKKWKKEDVIKQ--------------------L 336
                             .            .   .  :

newriceCDS701homologue      QVSDTDPYEAAVVSLPSAVKELLLRILRLRIGTRYDVSNLYIRSLLVSIL 441
rice                        EVSDTDPYEAAVVSLPSAVKELLLRILRLRIGTRYD-------------- 169
maize                       VADAADPYESAVNSLPSAVKEVLLRILRLRIGTRYD-------------- 199
CDS701Proteinprediction     EMDYSDPYEAAVVALPVVVKERLVRILRLGIATRYD-------------- 354
BAC83046.1                  KKIGIENIDSVTLKSDSNNPVCNRGFAFLELETSRDAR------------ 374
                                 :  ::..            :  * : *  *
```

FIGURE 1

```
newriceCDS701homologue      LFQIDIHCIRSLNELPEKAAVAVLNQCSQFLISGADKHNKGDYFASLIAK 491
rice                        ---IDIHCIRSLNELPEKAAVAVLN---QFLISGADKHNKGDYFASLIAK 213
maize                       ---IDIHCVKSLDELPESSALAVLN---QFLISGGDKHNKGDYFASLVAK 243
CDS701Proteinprediction     ---IDVESLTSLKILPQSAAISILD---QFMLSGADMQNKGGYLASLISK 398
BAC83046.1                  ---MAYKKLSQKNAFGKGLNIRVAWAEPLNDPDEKDMQVKSIFVDGIPTS 421
                               :  . : . . : :   : :         .  * : *. :. .: :.

newriceCDS701homologue      -------ETFSSALRLQGSTYLPRNPEIQNKRFPH---------SSRYSS 525
rice                        YQA----ETFSSALRLQGSTYLPRNPGIQNKRFPHQDYEYTASGSSRYSS 259
maize                       HQA----ETFGLTHALHGTTYLSRNPEMHSKRYPHEDYDFVTPRSSRYDS 289
CDS701Proteinprediction     QVEKLGPKQFDSRSRIEDVGLRVPEPDRFSTRVRLPDLDSYASRVPLPMP 448
BAC83046.1                  WDHAQLKEIFKKHGKIESVVLSRDMPSAKRRDFAFINYITREAAISCLES 471
                                   : *     :..       *                   .   .

newriceCDS701homologue      LGDYPS---SSYVDDPASSQSRNRRYDEYRPDLVRYPDSRSRQEEIVRIE 572
rice                        LGDYPS---SSYVDDPASSQSRNRRYDEYRPDLVRYPDSRSRQEEIVRIE 306
maize                       SAHHPS---TYYEDDPPVSESRVRRYAEERSTIVRSPEPRPRYDETD-IR 335
CDS701Proteinprediction     RTDVYTSHYSAYLDPHLSGRMTAKRMEEASSHLQATSLLSSRVATRMEEA 498
BAC83046.1                  FDKEEFSKNGSKVNIKVSLAKPAQQSKQTKEDHKSSITGEGKMKTSKIRY 521
                              .          :         ::  :             :

newriceCDS701homologue      RYPEPRFAHEPRQDTGRHLDLGYVQERNSNIERSAQVAFSSREGGYLSAS 622
rice                        RYPEPRFAHEPRQDTGRHLDLGYVQERNSNIERSAQVAFSSREGGYLSAS 356
maize                       INPEPRLPYESRHNAEKHLDRRYIQEHSSNIERPAEEALLSRERRFLPAA 385
CDS701Proteinprediction     GSTLQSLLSGGVTTRRMEEASPILQATLLPSGRVSRMDEASPNLQATWSP 548
BAC83046.1                  PVQDYTHIYSGEKRPFSTLGDPYYPLRGHSCRRHEGSTYTTAASSYGALP 571
                                                       *       :       .

newriceCDS701homologue      RYNTN-IVPEFSSRSSAEYSTARQQVRFDPFTGEPYKFDPYTGEPIRPES 671
rice                        RYNTN-IVPEFSSRSSAEYSTARQQVRFDPFTGEPYKFDPYTGEPIRPES 405
maize                       GYMPNPGGSDFRSRSPAEYSAQRQQMRFDPFTGEPYKFXPFTGEPIRPDP 435
CDS701Proteinprediction     SPTNDRIGLHSHITATADHQHTRPRIRFDPFTGEPYKFDPFTGEPIVPKS 598
BAC83046.1                  PATAESSLPHYHDSN----------RYPPHLGEAIKFSPTSAVLSKQAW 610
                                :    .               *: *. **. ** * :.

newriceCDS701homologue      NPR--RSGSLY-- 680
rice                        NPR--RSGSLY-- 414
maize                       NPAPLRKPVIXSE 448
CDS701Proteinprediction     SSH---HRSLY-- 606
BAC83046.1                  QKM---------- 613
```

FIGURE 1 (continued)

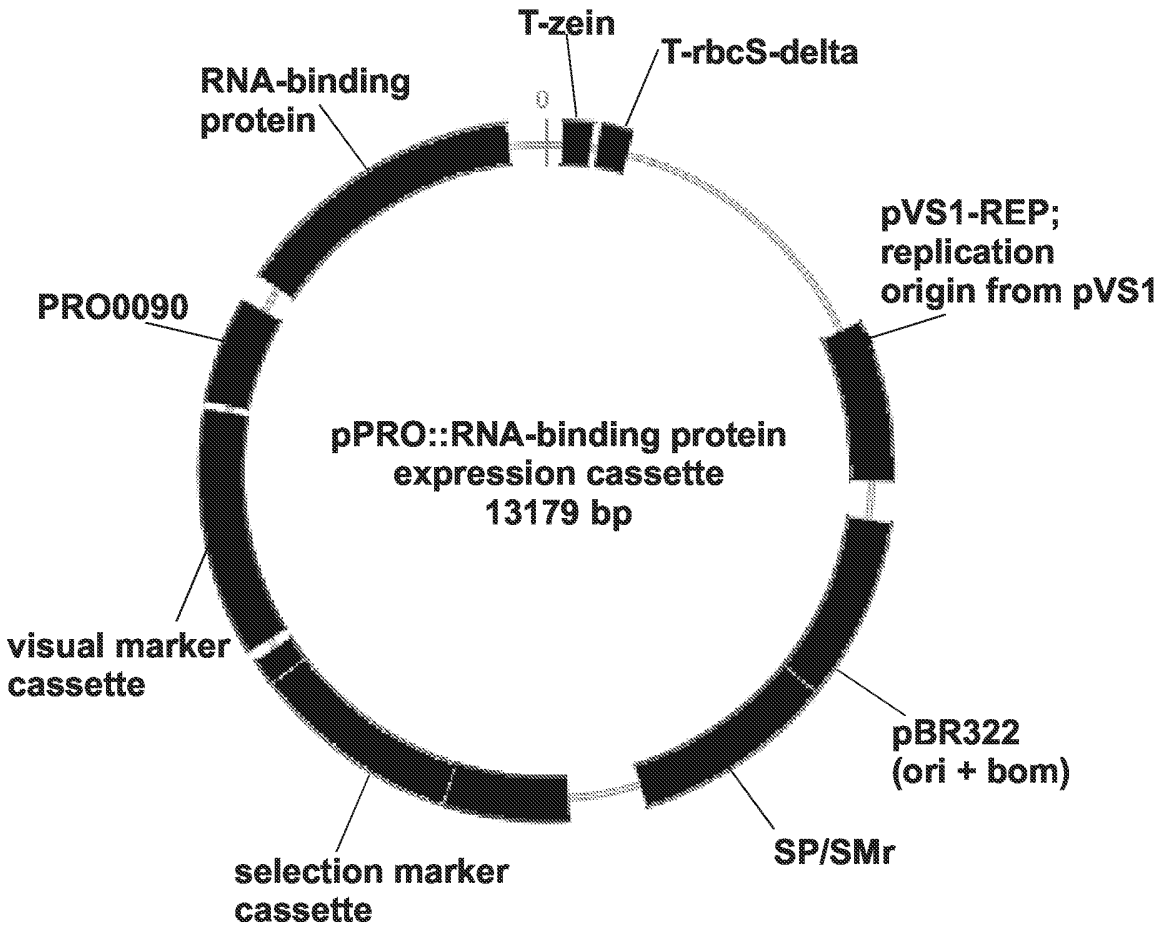

T-zein
T-rbcS-delta
RNA-binding protein
0
pVS1-REP; replication origin from pVS1
PRO0090
pPRO::RNA-binding protein expression cassette 13179 bp
visual marker cassette
pBR322 (ori + bom)
selection marker cassette
SP/SMr

FIGURE 2

1 50
Translation of AK067725-Os-RBP1 (1) ----------MEADSGKLFVGGISWETDEDRLREYFSRFGEVTEAVIMRDR
Translation of AK070544-Os-RBP1 (1) ----------MEADAGKLFIGGISWDTNEDRLREYFTKYGEVVEAVIMRDR
Translation of NM_196957-Os-RBP1 (1) ----------MESDQGKLFIGGISWETTEEKLRDHFAAYGDVSQAAVMRDK
NP_176143-At-RBP1 (1) ----------MDYDRYKLFVGGIAKETSEEALKQYFSRYGAVLEAVVAKEK
NP_567753-At-RBP1 (1) MNPEEQKMESASDLGKLFIGGISWDTDEERLQEYFGKYGDLVEAVIMRDR
NP_974937-At-RBP1 (1) ----------MESDLGKLFIGGISWDTDEERLRDYFSNYGDVVEAVIMRDR
NP_193166-At-RBP1 (1) ----------MDSDQGKLFVGGISWETDEDKLREHFTNYGEVSQAIVMRDK
NP_850539-At-RBP1 (1) ----------MQSDNGKLFIGGISWDTNEERLKEYFSSFGEVIEAVILKDR
NP_180899-At-RBP1 (1) ----------MESDQGKLFIGGISWDTDEMLLREYFSNFGEVLQVTVMREK
NP_974899-At-RBP1 (1) ----------
51 100
Translation of AK067725-Os-RBP1 (42) NTGRARGFGFVVFTDAGVAERVTMDKHMIDGRMVEAKKAVPRDDQSITSK
Translation of AK070544-Os-RBP1 (42) ATGRARGFGFIVFADPAVAERVIMEKHMIDGRMVEAKKAVPRDDQHALSK
Translation of NM_196957-Os-RBP1 (42) LTGRPRGFGFVVFSDPSSVDAALVDPHTIDGRTVDVKRALSREECQAAKA
NP_176143-At-RBP1 (42) VTGKPRGFGFVRFANDCDVVKALRDTHFILGKPVDVRKAIRKHELYQQPF
NP_567753-At-RBP1 (51) TTGRARGFGFIVFADPSVAERVIMDKHIIDGRTVEAKKAVPRDDQQVLKR
NP_974937-At-RBP1 (42) ATGRARGFGFIVFADPCVSERVIMDKHIIDGRTVEAKKAVPRDDQQVLKR
NP_193166-At-RBP1 (42) LTGRPRGFGFVIFSDPSVLDRVLQEKHSIDTREVDVKRAMSREECQVSGR
NP_850539-At-RBP1 (42) TTGRARGFGFVVFADPAVAELVITEKHNIDGRLVEAKKAVPRDDQNMVNR
NP_180899-At-RBP1 (42) ATGRPRGFGFVAFSDPAVIDRVLQDKHHIDNRDVDVKRAMSREEQSPAGR
NP_974899-At-RBP1 (1) ----------MVEAKKAVPRDDHVVFNK
101 150
Translation of AK067725-Os-RBP1 (92) ANGS----------SIGSPG----------PGRTRKIFVGGLASNVTE
Translation of AK070544-Os-RBP1 (92) SGGS----------AHGSPG----------PSRTKKIFVGGLASTVTE
Translation of NM_196957-Os-RBP1 (92) ANPSAGGRHASGGGGGGGGAGGGGGGGGDAGGARTKKIFVGGLPSNLTE
NP_176143-At-RBP1 (92) SMQFLERKVQQMNGGLREMS----------SNGVTSRTKKIFVGGLSSNTTE
NP_567753-At-RBP1 (101) HASPMHLISPSHGGNGG----------GARTKKIFVGGLPSSITE
NP_974937-At-RBP1 (92) HASPIHLMSPVHGGGG----------RTKKIFVGGLPSSITE
NP_193166-At-RBP1 (92) TGNLNTSRSSGG----------DAYNKTKKIFVGGLPPTLTD
NP_850539-At-RBP1 (92) SNSS----------IQGSPGG----------PGRTRKLFVGGLPSSVTE
NP_180899-At-RBP1 (92) SGTFNASRNFDS----------GANVRTKKIFVGGLPPALTS
NP_974899-At-RBP1 (19) SNSS----------LQGSPG----------PSNSKKIFVGGLASSVTE

FIGURE 3

```
                                            151                                              200
Translation of AK067725-Os-RBP1    (120) VEFRRYFEQFGVITDVVVMYDHNTQRPRGFGFITYDSEDAVDKALHKNFH
Translation of AK070544-Os-RBP1    (120) ADFRKYFEQFGTITDVVVMYDHNTQRPRGFGFITYDSEDAVDKALFKTFH
Translation of NM_196957-Os-RBP1   (142) DEFRQYFQTYGVVTDVVVMYDQNTQRPRGFGFITFDAEDAVDRVLHKTFH
               NP_176143-At-RBP1   (134) EEFKSYFERFGRTTDVVVMHDGVTNRPRGFGFVTYDSEDSVEVVMQSNFH
               NP_567753-At-RBP1   (136) AEFKNYFDQFGTIADVVVMYDHNTQRPRGFGFITFDSEESVDMVLHKTFH
               NP_974937-At-RBP1   (124) EEFKNYFDQFGTIADVVVMYDHNTQRPRGFGFITFDSDDAVDRVLHKTFH
               NP_193166-At-RBP1   (124) EEFRQYFEVYGPVTDVAIMYDQATNRPRGFGFVSFDSEDAVDSVLHKTFH
               NP_850539-At-RBP1   (122) SDFKTYFEQFGTTTDVVVMYDHNTQRPRGFGFITYDSEEAVEKVLLKTFH
               NP_180899-At-RBP1   (124) DEFRAYFETYGPVSDAVIMIDQTTQRPRGFGFVSFDSEDSVDLVLHKTFH
               NP_974899-At-RBP1    (47) AEFKKYFAQFGMITDVVVMYDHRTQRPRGFGFISYDSEEAVDKVLQKTFH

                                            201                                              250
Translation of AK067725-Os-RBP1    (170) ELNGKMVEVKRAVPKEQSPGPA--ARSPAGG--QNYAMSRVHSFLNGFNQ
Translation of AK070544-Os-RBP1    (170) ELNGKMVEVKRAVPKELSPGPS--MRSPVGG--FNYAVNRANNFLNGYTQ
Translation of NM_196957-Os-RBP1   (192) DLSGKMVEVKRALPREANPGSGSGGRSMGGGGGGYQSNNGPNSNSGGYDS
               NP_176143-At-RBP1   (184) ELSDKRVEVKRAIPKE------------------G--IQSNNGNAVNIPP
               NP_567753-At-RBP1   (186) ELNGKMVEVKRAVPKELSSTTP--NRSPLIGYGNNYGVVPNRSSANSYFN
               NP_974937-At-RBP1   (174) ELNGKLVEVKRAVPKEISPVSN--IRSPLAS-GVNYGGGSNRMPANSYFN
               NP_193166-At-RBP1   (174) DLSGKQVEVKRALPKDANPGGG-GRSMGGGGSGGYQGYGGNESS---YDG
               NP_850539-At-RBP1   (172) ELNGKMVEVKRAVPKELSPGPS--RSPLGAG--YSYGVNRVNNLLNGYAQ
               NP_180899-At-RBP1   (174) DLNGKQVEVKRALPKDANPGIASGGGRGSGGAGGFPGYGGSGGSG--YEG
               NP_974899-At-RBP1    (97) ELNGKMVEVKLAVPKDMALNTMRNQMN-----VNSFGTSRISSLLNEYTQ

                                            251                                              300
Translation of AK067725-Os-RBP1    (216) GYNPNPIGGYGMRVDG-RYGLLT-GARNGFSSFGPGYGMGMNSESGMNAN
Translation of AK070544-Os-RBP1    (216) GYNPSPVGGYGMRMDA-RFGLLS-GGRSSYPSFGGGYGVGMNFDPGMNPA
 ranslation of NM_196957-Os-RBP1   (242) RGDASRYGQAQQGSGG---YPGY--GAGGYGAGTVGYGYGHANPGTAYGN
               NP_176143-At-RBP1   (214) SYSSFQATPYVPEQNG-------YGMVLQFPPPVFGYHHNVQAVQYPYGY
               NP_567753-At-RBP1   (234) SFPPG-YNNNNLGSAG-RFSPIG-SGRNAFSSFGLGLNQELNLNSN----
               NP_974937-At-RBP1   (221) NFAPGPGFYNSLGPVGRRFSPVIGSGRNAVSAFGLGLNHDLSLNLNPSCD
               NP_193166-At-RBP1   (220) RMDSNRFLQHQSVGNG---LPSY--GSSGYGA---GYGNG--SNGAGYG-
               NP_850539-At-RBP1   (218) GFNPAAVGGYGLRMDG-RFSPVG-AGRSGFANYSSGYGMNVNFDQGLPTG
               NP_180899-At-RBP1   (222) RVDSNRYMQPQNTGSG---YPPY--GGSGYGT---GYGYG--SNGVGYG-
               NP_974899-At-RBP1   (142) GFSPSPISGYGVKPEVRYSPAVG--NRGGFSPFGHGYGIELNFEPNQTQN
```

FIGURE 3(continued)

```
                                           301                                                350
Translation of AK067725-Os-RBP1    (264) FGANSSFVNNSN-GRQIGSFYNGSS-NRLGSPIGYVGLN-DDSGSLLSSM
Translation of AK070544-Os-RBP1    (264) IGGSSSFNNSLQYGRQLNPYYSGNS-GRYNSNVSYGGVN-DSTGSVFNSL
Translation of NM_196957-Os-RBP1   (287) YGAGGFGGVPAGYG----GHYGNPN----APGSGYQGGP--------PGA
              NP_176143-At-RBP1    (257) QFTAQVANVSWN--------------NPIMQPTGFYCAP-----------
              NP_567753-At-RBP1    (277) FDGNTLGYSRIP----GNQYFNSASPNRYNSPIGYNRGD-----SAYNPS
              NP_974937-At-RBP1    (271) GTSSTFGYNRIP----SNPYFNGASPNRYTSPIGHNRTE-----SPYNSN
              NP_193166-At-RBP1    (259) -AYGGYTGSAGGYGAGATAGYGATN----IPGAGYGSST--------GVA
              NP_850539-At-RBP1    (266) FTGGTNYNGNVDYGRGMSPYYIGNT-NRFGPAVGYEGGNGGGNSSFFSSV
              NP_180899-At-RBP1    (261) -GFGGYGNPAG-------APYGNPS----VPGAGFGS------------G
              NP_974899--At-RBP1   (190) YGSGSSGGFGRPFSPGYAASLGRFGSQMESGGASVGNGS-----VLNAAP

                                           351                                                400
Translation of AK067725-Os-RBP1    (311) SRNVWGNEN-LNYPNNPTNMSS--FAPSGTGG----QMGITSDGINWGG-
Translation of AK070544-Os-RBP1    (312) ARNLWGNSG-LSYSSNSASSNS--FMSSANGG----LGGIGNNNVNWGNP
Translation of NM_196957-Os-RBP1   (321) NRGPWGGQAPSGYG---------------TGS---------------YGG
              NP_176143-At-RBP1    (282) PHPTPPPTNNLG--------------------------------------
              NP_567753-At-RBP1    (318) NRDLWGNRS-------------------DSSG------------------
              NP_974937-At-RBP1    (312) NRDLWGNRT-------------------DTAG------------------
              NP_193166-At-RBP1    (296) PRNSWDTPASSGYG---------------NPG---------------YGS
              NP_850539-At-RBP1    (315) TRNLWGNNGGLNYNNNNTNSNSNTYMGGSSSGNNTLSGPFGNSGVNWGA-
              NP_180899-At-RBP1    (287) PRSSWGAQAPSGYG---------------NVG---------------YGN
              NP_974899-At-RBP1    (235) KNHLWGNGG-LGYMS---------------NS------------------

                                           401                                                450
Translation of AK067725-Os-RBP1    (353) PTPGHGMGNISSLGLANLGRGAG--DSFGLPSGSYGRSNATGTIGEPFS-
Translation of AK070544-Os-RBP1    (355) PVPAQGANAGPGYGSGNFGYGSSE-TNFGLGTNAYGR-NAGSGVVNTFN-
Translation of NM_196957-Os-RBP1   (341) NAGYAAWNNSSAGGNAPTSQAAGAGTGYGSQGYGYGG-YGGDASYGNHGG
              NP_176143-At-RBP1    (294) --YIQYMNGFDLSGTNISGYNPLAWPVTGDAAGALIHQFVDLKLDVHSQA
              NP_567753-At-RBP1    (331) --PGWNLGVSVGNNRGNWGLSS-----VVSDNNGYGRSYGAGSGLSGLSF
              NP_974937-At-RBP1    (325) --PGWNLNVSNGNNRGNWGLPSS--SAVSNDNNGFGRNYGTSSGLS----
              NP_193166-At-RBP1    (316) GAAHSGYG---VPGAAPPTQS---PSGYSNQGYGYGGYSGSDSGYGNQAA
              NP_850539-At-RBP1    (364) --PGGGNNAVSNEN-VKFGYGGNGESGFGLGTGGYAARNPGANKAAPSSS
              NP_180899-At-RBP1    (307) AAPWGGSG---GPGSAVMGQAG-ASAGYGSQGYGYG---GNDSSYGTPSA
              NP_974899-At-RBP1    (251) --PISRSSFSGNSGMSSLGSIGDNWGTVARARSSYHGERGGVGLEAMRG-
```

FIGURE 3(continued)

```
                                          451                                                500
Translation of AK067725-Os-RBP1   (400) -APPNAYEVNNADTYGSS------SIYGDSTWRFTSSEID--------MP
Translation of AK070544-Os-RBP1   (402) -QSTNGYGRNFGDSSGGGGGGGGGSIYGDTTWRSGSSELDG-------TS
Translation of NM_196957-Os-RBP1  (390) YGGYGGRGDGAGNPAAGG------GSGYGAGYGSGNGGSGYP------NA
              NP_176143-At-RBP1   (342) HQRMNGGNMGIPLQNGTYI-------------------------------
              NP_567753-At-RBP1   (374) AGNTNGFDGSIGELYRGS------SVYSDSTWQQSMPHHQSSNELDGLSR
              NP_974937-At-RBP1   (367) SSPFNGFEGSIGELYRGG------SVYSDSTWQQQLPSQSSHELDNLSR
              NP_193166-At-RBP1   (360) YGVVGGRPSGGGSNNPGS------GGYMGGGYGDGS-------------W
              NP_850539-At-RBP1   (411) FSSASATNNTGYDTAGLAEFYGNGAVYSDPTWRSPTPETEG-------PA
              NP_180899-At-RBP1   (350) YGAVGGR-SGNMPNNHGG------GGYADALDGSGG-------------Y
              NP_974899-At-RBP1   (298) -VHVGGYSSGSSILEADS-------LYSDSMWLSLP--AK---------A

                                          501                                                550
Translation of AK067725-Os-RBP1   (435) PFGNDLGN--VDPDIKSNIP-ASYMGNYTVNNNQTSRGITS---------
Translation of AK070544-Os_RBP1   (444) PFGYGLGN--AASDVTAKNS-AGYMGH-----------------------
Translation of NM_196957-Os_RBP1  (428) WADPSQGGGFGASVNGVSEGQSNYGSGYGGVQPRVAQ-------------
              NP_176143-At-RBP1   (361) --------------------------------------------------
              NP_567753-At-RBP1   (418) SYGFGIDN--VGSDPSANAS-EGYSGNYNVGNRQTHRGIEA---------
              NP_974937-At-RBP1   (411) AYGYDIDN--VGSDPSANDP-ETYNGSYNVGNRQTNRGNKKIHLNKTCNL
              NP_193166-At-RBP1   (391) RSDPSQ----------------GYGGGYNDGQGRQGQ-------------
              NP_850539-At-RBP1   (454) PFSYGIGGGVPSSDVSARSSSPGYVGSYSVNKRQPNRGEPSR--------
              NP_180899-At-RBP1   (380) GNHQGNNG------------QAGYGGGYGSGRQAQQQ-------------
              NP_974899-At-RBP1   (329) EEGLGMGP----LDFMS--R-----GPAGYINRQPNGGIAA---------

                                          551
Translation of AK067725-Os-RBP1   (473) ---
Translation of AK070544-Os-RBP1   (468) ---
Translation of NM_196957-Os-RBP1  (465) ---
              NP_176143-At-RBP1   (361) ---
              NP_567753-At-RBP1   (456) ---
              NP_974937-At-RBP1   (458) DTF
              NP_193166-At-RBP1   (412) ---
              NP_850539-At-RBP1   (496) ---
              NP_180899-At-RBP1   (405) ---
              NP_974899-At-RBP1   (359) ---
```

FIGURE 3(continued)

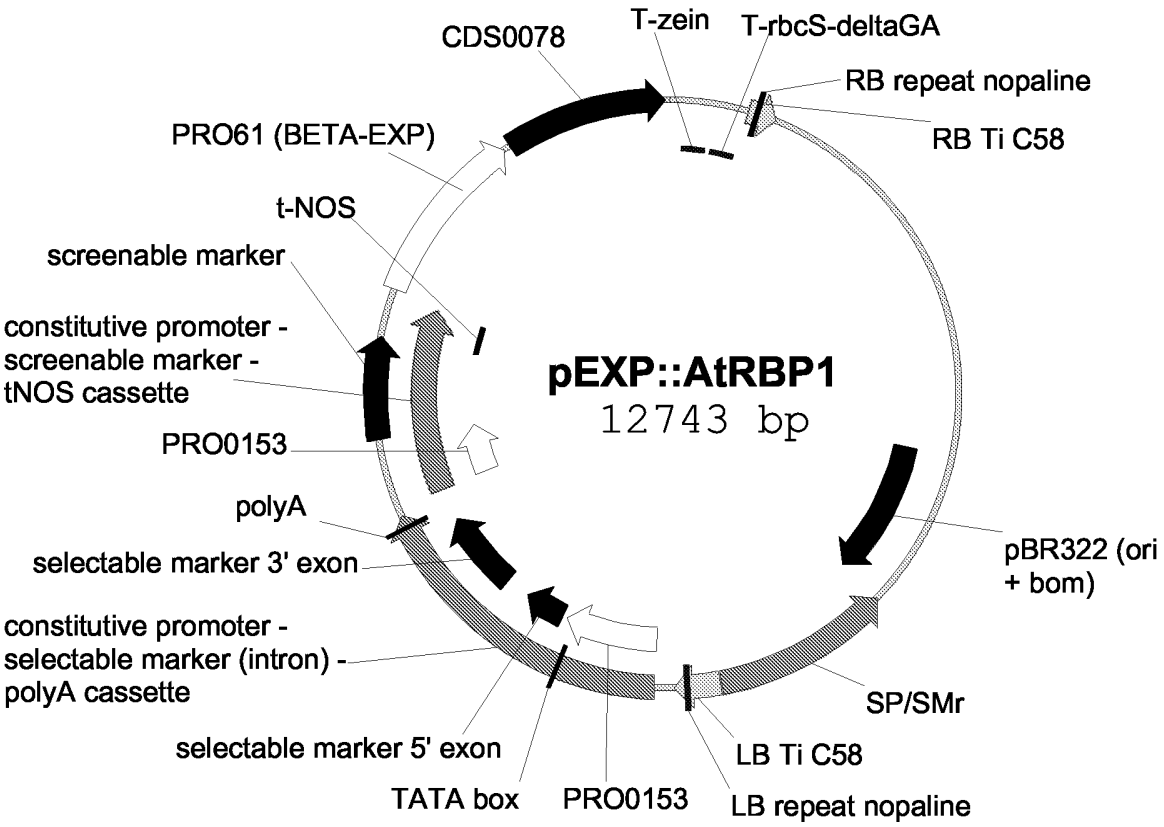
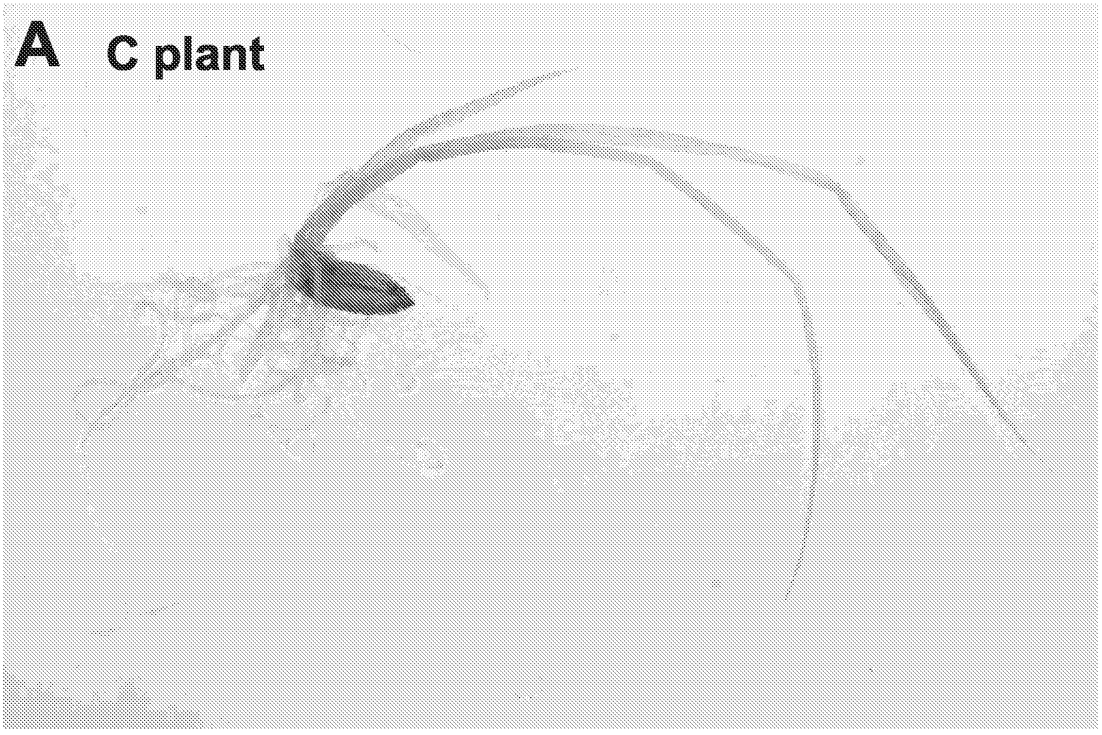

CDS0078
T-zein
T-rbcS-deltaGA
RB repeat nopaline
RB Ti C58
PRO61 (BETA-EXP)
t-NOS
screenable marker
constitutive promoter -
screenable marker -
tNOS cassette
PRO0153
polyA
selectable marker 3' exon
constitutive promoter -
selectable marker (intron) -
polyA cassette
selectable marker 5' exon
TATA box
PRO0153
LB repeat nopaline
LB Ti C58
SP/SMr
pBR322 (ori
+ bom)
pEXP::AtRBP1
12743 bp

FIGURE 4



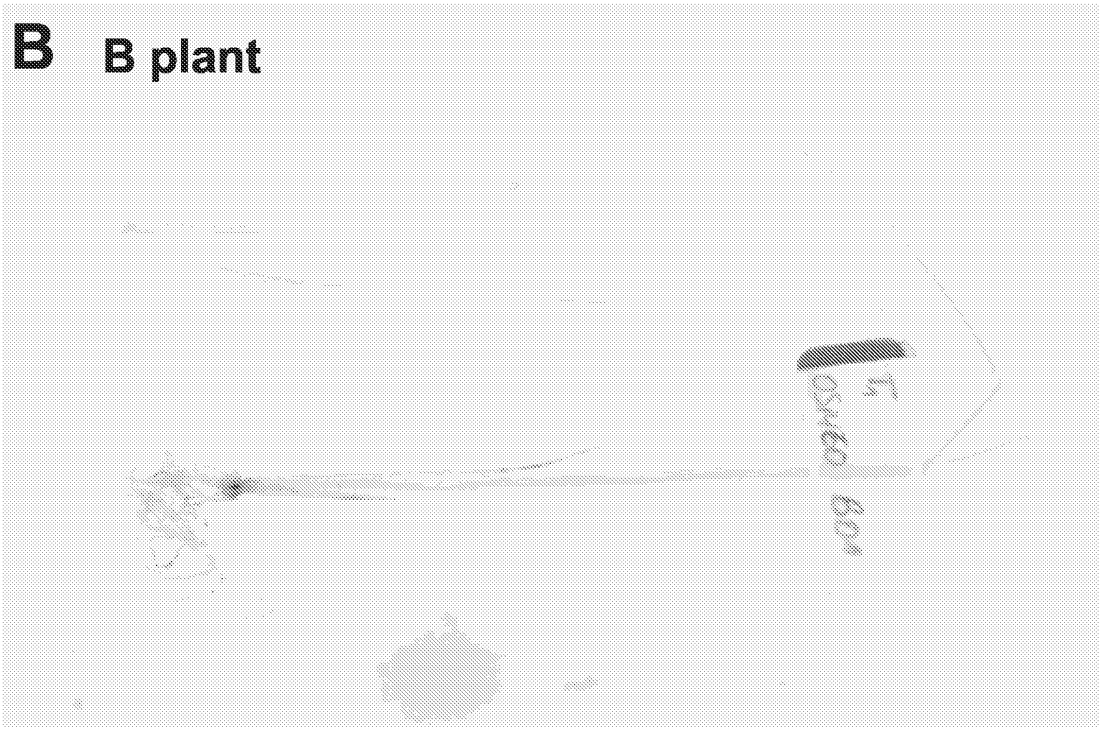


FIGURE 5

seq id no 01: *Nicotiana tabacum* RNA-binding protein DNA

CCACGCGTCCGCTTAGGGTTCCAAATTGCTCTAAATTCCCGCGGATTGAGAGTTCATTGGAGAC
TTCCATTGTTCCCAGCGGCTAAGATGAGCCGGTTGATTGAGCATCACCTAGCAAATAATAAACA
GGACATGAAAGGGACAGAGGTTTTTGTTGGTGGTTTGGCCCGTACTACTACTGAAAGCAAAATT
CATGAGGTATTTTCTTCATGTGGTGAGATTGTGGAAATACGGTTGATAAAAGACCAGACAGGCG
TTCCTAAGGGGTTTTGCTTTGTACGATTTGCAACAAAATATGCTGCTGACAAAGCTCTGAAGGA
AAAATCTGGATATGTGCTGGATGGGAAGAAACTCGGGGTTCGCCCCTCAGTTGAGCAGGACACT
TTATTTCTTGGAAATCTTAACAAAGGTTGGGGTGCGGAGGAATTTGAGAGTATTGTGCGCCAGG
TTTTTCCAGATGTTGTATCTGTTGATCTTGCACTTCTTGGAGATGTCCAACCTGGTCAGAAGCA
ACGGAATCGGGGTTTTGCTTTCGTGAAATTCCCATCTCATGCTGCTGCGGCTCGTGCTTTTCGG
GTAGGCTCCCAATCTGATTTTCTCATTGATGGCAAGTTACATCCATCTGTACAGTGGGCTGAGG
AACCTGATCCCAATGAACTTGCTCAGATCAAAGCAGCCTTCGTTAGAAATGTACCTCCTGGTGC
TGATGAAGATTACTTGAAGAAGCTCTTTCAGCCCTTTGGCAATGTAGAGAGGATAGCTCTATCC
AGGAAAGGTAGCTCCACCATTGGATTCGTTTACTTCGATAAGCGATCTGATCTTGACAATGCTA
TTATGGCGTTGAATGAGAAAACTGTACAAGGGCCAATGGGAGGTCCCTCATGCAAGCTTCAGGT
CGAAGTTGCTAGGCCAATGGACAAGAACAGGAAACGAGGTCGTGAGGATCCAAACATGTCCAGT
ACCATTGAGAGTCATTCCAAGCTTTTGAAGGATGATCCAGATGTTGAGATGATTAGGGCTCCTA
AATCAACTGCTCAACTGGAGATGGATTATTCGGATCCTTATGAAGCTGCTGTAGTTGCATTACC
TGTGGTTGTCAAGGAGCGTTTAGTTCGGATCTTGCGGCTTGGTATTGCTACTAGATATGATATA
GATGTTGAAAGTTTAACCAGTCTTAAGATATTGCCCCAGTCAGCTGCCATATCTATTCTTGACC
AGTTCATGTTGTCTGGAGCTGATATGCAGAACAAGGGAGGATATCTAGCTTCATTAATTTCTAA
GCAGGTTGAAAAACTGGGACCGAAACAATTCGATAGTAGGTCAAGGATAGAAGATGTTGGCTTG
AGGGTGCCAGAACCAGACAGGTTCTCTACAAGAGTTCGTTTGCCAGATCTAGATTCATATGCCT
CACGAGTACCCTTGCCCATGCCTAGGACTGATGTTTACACATCTCACTATTCAGCGTATTTAGA
TCCCCATCTGTCTGGTCGGATGACAGCAAAGAGGATGGAGGAAGCAAGTTCCCATTTGCAGGCG
ACTTCACTTCTGTCTAGTCGGGTGGCAACGAGGATGGAGGAGGCAGGTTCCACTTTGCAGTCGC
TCCTATCTGGTGGGGTGACGACAAGAAGGATGGAGGAAGCAAGTCCGATTTTGCAGGCAACACT
CCTTCCATCTGGTCGGGTATCAAGGATGGATGAAGCAAGTCCCAATTTGCAGGCAACATGGAGC
CCTTCTCCTACTAATGACAGAATTGGACTTCATTCACACATTACCGCAACTGCTGATCATCAAC
ATACTCGACCACGGATCAGGTTTGATCCCTTCACTGGTGAGCCATACAAATTTGACCCCTTCAC
TGGCGAGCCAATTGTTCCCAAGAGCTCAAGTCATCATCGAAGCCTGTACTGAACGTTCTGAGCA
TTCTAATTTACAAATGGCTTATTGCCAAACCTATGTAACATAATGATGCGTATTTTTGTTCATC
CGCAGCTGTAAAATAGTAGCTGTTAGCAGGATTATTTGGTTATGTTTCTCATTGACTTCATTGA
TTGCGAAGGTGCATTTGGAATCTCGGCAATCACAATTTATAGCCGGTGCA

seq id no 02: *Nicotiana tabacum* RNA-binding protein

MSRLIEHHLANNKQDMKGTEVFVGGLARTTTESKIHEVFSSCGEIVEIRLIKDQTGVPKGFCFV
RFATKYAADKALKEKSGYVLDGKKLGVRPSVEQDTLFLGNLNKGWGAEEFESIVRQVFPDVVSV
DLALLGDVQPGQKQRNRGFAFVKFPSHAAAARAFRVGSQSDFLIDGKLHPSVQWAEEPDPNELA
QIKAAFVRNVPPGADEDYLKKLFQPFGNVERIALSRKGSSTIGFVYFDKRSDLDNAIMALNEKT
VQGPMGGPSCKLQVEVARPMDKNRKRGREDPNMSSTIESHSKLLKDDPDVEMIRAPKSTAQLEM
DYSDPYEAAVVALPVVVKERLVRILRLGIATRYDIDVESLTSLKILPQSAAISILDQFMLSGAD
MQNKGGYLASLISKQVEKLGPKQFDSRSRIEDVGLRVPEPDRFSTRVRLPDLDSYASRVPLPMP
RTDVYTSHYSAYLDPHLSGRMTAKRMEEASSHLQATSLLSSRVATRMEEAGSTLQSLLSGGVTT
RRMEEASPILQATLLPSGRVSRMDEASPNLQATWSPSPTNDRIGLHSHITATADHQHTRPRIRF
DPFTGEPYKFDPFTGEPIVPKSSSHHRSLY

FIGURE 6

seq id no 03: *Oryza sativa* RNA-binding protein DNA

ATGGTGCGTGCTCGAGACTCAATCCGCGAAATCCTCCCTGTTTTTTCGATTCAATCCGCCCTGG
GGACGGCGGATTCGGCGCCGGCGATCCGGCCGGTCGCCGCCGCGTCCGATTTGGTGCGGATTTC
GTCGGAGAAATCGCGTCTTGACCTTCCTGTGCCTCTTTTTTTTTTTGTTGCTCGTGGGGGATTT
CAGGAGAAGAGGGGGGCGGCGTCGCATGGCGACTACGACGAGCAAGGTTATTGGATGGGTTTCT
TCTCTTTGATACCTCGAGCGAGTCTTGCGTTGCGTGGGTGAAAGGCGCCGAGGTGTTCGTCGGC
GGGTTGCCGCGGTCGGTGACGGAGCGGGCGCTCCGAGAGGTTGGTGTTCTTCCGAGAGGTGTAA
TCTCAACAGGTATTTTCTCCTTGTGGAGAGATTGTTGATTTGCGGATAATGAAAGATCAGAATG
GCATTTCAAAGTGGTTCTCTGCCAGCTTCAAGGAAAGAGACTTGCTGTTGATCTTTCGTTGGAT
CAAGATACACTCTTCTTTGGGAATCTTTGCAAAGGTAGTCAGACTGGGGCATCGAAGAATTTGA
AGAATTGATTCGCAAGGTAAGACCTGTAGGTTGACCTTGCAATGGCTCGAAACCATGACTCTTC
AGTTGGGAAAAGACGTCTAAATCGAGGCTTTGCATTTGTGCGATTTTCTTCTCATGCAGTAAGT
GTTGACATGATAACCCTTTTCTGCCAATTTTCTTTTTTGCAGGTGTCTGATACGGACCCCTATG
AAGCAGCTGTTGTTTCACTACCTTCAGCCGTCAAGGAACTCCTACTTCGTATTCTACGTCTTAG
AATTGGCACTCGATATGATGTAAGTAATCTGTACATAAGGTCTCTACTTGTGCAGCTCCAGGTC
ATCTGCTGAATACTCTACTGCTCGCCAACAAGTAAGGTTTGATCCATTCACAGGGGAACCATAC
AAGTTTGATCCCTACACCGGTGAACCCATCAGGCCAGAATCGAACCCACGTCGCTCAGGAAGCT
TATACTGACTTTGATTGATTGAAGCAACAGTTTGGATATGGTAGATTAGATTTACATCCCTGAA
CCAAAAGGACCATAT

seq id no 04: *Oryza sativa* RNA-binding protein

MVRARDSIREILPVFSIQSALGTADSAPAIRPVAAASDLVRISSEKSRLDLPVPLFFFVARGGF
QEKRGAASHGDYDEQGYWMGFFSLIPRASLALRGRRVKGAEVFVGGLPRSVTERALREVGVLPR
SQQVFSPCGEIVDLRIMKDQNGISKVLCQLQGKRLAVDLSLDQDTLFFGNLCKGSDWGIEEFEE
LIRKVRPVVDLAMARNHDSSVGKRRLNRGFAFVRFSSHAVSQVKTAFVGNLPANVTEEYLRKLF
EHCGEVCYAVVRVAVSRKGQYPVGFVHFASRTWKELDNAIKEMDGETVRGPDRGATFRIQVSVA
RPVVENDKKRIREEVKTRRSNVSTDKPDHSYGRRGHDSYDRQAKAPRLYNEVLHTNDKVDMITL
FCQFSFLQVSDTDPYEAAVVSLPSAVKELLLRILRLRIGTRYDVSNLYIRSLLVSILLFQIDIH
CIRSLNELPEKAAVAVLNQCSQFLISGADKHNKGDYFASLIAKETFSSALRLQGSTYLPRNPEI
QNKRFPHSSRYSSLGDYPSSSYVDDPASSQSRNRRYDEYRPDLVRYPDSRSRQEEIVRIERYPE
PRFAHEPRQDTGRHLDLGYVQERNSNIERSAQVAFSSREGGYLSASRYNTNIVPEFSSRSSAEY
STARQQVRFDPFTGEPYKFDPYTGEPIRPESNPRRSGSLY

seq id no 05: *Zea mays* RNA-binding protein DNA (AY105295)

TCTAGCTGTGTTCTTGTGGCTGTGAAATTATATCTCCCATGCTGATACTTGATTCCCTTATCTT
TGCTTCATTACTACACCACAGTAATTTGGATCTGCCATTATGTTACTATGTAACTCTCATTTGA
TATCAATCACAGCTGCCACATACAAAATACAAGTATGTTTATCTAGATAAGATCTTGATTCATC
AATCACCACTGATCTGAGTTTTCGCCACTGCGATGCGAGGAAAAGACAGATATCTAATAACATC
TTGGTGAAGATGTTCTTAGGTCCTTTGCTTTCTCTTCAAGTCAGCTTCCTTTGATTTCATTCCT
CAAACTATCAATCACAGGCTGCAGCACGTGTAATCCGCATCGGTTCAAGAACAGATTTCATGCT
TGGTGATATTTTGCATCCTGCGATAAATTGGGCTGATAAAGAGTCTCATCTGGATCCTGATGAA
ATGGCCAAGATGAAGTCTGCTTTTATTGGTAACCTGCCAGAAGATGTTAATGAGGAGTACTTGA
GAAAGCTTTTTGGACAGTTCGGTGAGGTAGTACGGGTTGCTATCTCAAGAAAAGGACAATGTCC
AGTTGCTTTTGTTCACTTCGCCAAACGTTCAGAGCTTGAGAATGCTATAGAAGAAATGGATGGT

FIGURE 6

AAAACGGTGAGAGGACCTGGTCGAGGGCCGTCTTTCAAGATCCAGGTGTCAGTTGCTCGACCTA
CGGCAGACAACGACAAGAAGCGATCTCGTGAAGAAGTGAGAACTAGAAGATCAAATGCATCAGG
AGATAGGCGAGATTATTCTCATGGAAGATATGGACACGATTCACTTGATCGTCAAGTGAAAGCT
CCAAGATTATCTAATTATGTGGCCGATGCTGCTGACCCCTATGAATCAGCTGTTAATTCATTAC
CTTCAGCTGTCAAGGAAGTCTTGCTTCGAATTCTACGTCTAAGAATTGGTACTCGATATGATAT
TGATATCCATTGTGTTAAAAGCCTTGATGAGCTTCCTGAGTCATCTGCTCTTGCTGTCCTTAAT
CAGTTTTTGATATCAGGTGGAGACAAACACAACAAAGGAGATTATTTTGCATCGTTGGTTGCTA
AGCACCAGGCTGAGACCTTTGGCTTAACACATGCATTACACGGTACCACTTATTTGTCAAGAAA
TCCGGAAATGCATAGCAAGCGATACCCACATGAAGATTATGATTTTGTGACACCCAGGAGCAGT
AGGTACGATTCGTCAGCCCATCATCCTTCAACATACTACGAAGACGATCCACCAGTGTCTGAGT
CAAGGGTTAGAAGATATGCTGAAGAAAGGTCCACCATTGTAAGAAGCCCAGAACCACGTCCGCG
ATATGACGAAACAGACATAAGAATAAACCCAGAACCAAGATTACCATATGAATCAAGACACAAC
GCCGAAAAGCATCTCGATCGAAGATACATACAAGAGCATAGTTCAAATATTGAAAGACCAGCTG
AAGAAGCTCTCCTTTCTAGGGAAAGGAGATTTCTGCCTGCTGCAGGGTACATGCCGAACCCAGG
CGGCTCGGATTTCCGCTCCAGGTCGCCCGCCGAATATTCAGCACAACGCCAACAAATGAGGTTT
GATCCATTCACAGGTGAACCTTACAAGTTTGNACCCTTCACAGGGGAGCCCATCAGGCCAGATC
CGAACCCAGCGCCGCTCAGGAAGCCTGTAATTGANTCAGAATAAGTTTGGAAGCCGANAATGCC
AGATTAAGAACCCTGAAANCAAAGCNAAGA

seq id no 06: *Zea mays* RNA-binding protein (AY105295)

GSRTDFMLGDILHPAINWADKESHLDPDEMAKMKSAFIGNLPEDVNEEYLRKLFGQFGEVVRVA
ISRKGQCPVAFVHFAKRSELENAIEEMDGKTVRGPGRGPSFKIQVSVARPTADNDKKRSREEVR
TRRSNASGDRRDYSHGRYGHDSLDRQVKAPRLSNYVADAADPYESAVNSLPSAVKEVLLRILRL
RIGTRYDIDIHCVKSLDELPESSALAVLNQFLISGGDKHNKGDYFASLVAKHQAETFGLTHALH
GTTYLSRNPEMHSKRYPHEDYDFVTPRSSRYDSSAHHPSTYYEDDPPVSESRVRRYAEERSTIV
RSPEPRPRYDETDIRINPEPRLPYESRHNAEKHLDRRYIQEHSSNIERPAEEALLSRERRFLPA
AGYMPNPGGSDFRSRSPAEYSAQRQQMRFDPFTGEPYKFXPFTGEPIRPDPNPAPLRKPVIXSE

seq id no 07: *Oryza sativa* RNA-binding protein DNA (AK059444)

ATCGATCACAGGCTGCAGCACGCGTACTTCGTATTGGTTCCAGAACAGATTTTCTGCTTGGTGG
ATTGCATCCTTCAATAAATTGGGCTGAGAAGGAGTCTCATGTAGATGAGGACGAAATGGCCAAG
GTTAAGACAGCTTTCGTTGGAAATTTACCAGCAAATGTTACAGAGGAGTATTTAAGAAAGCTTT
TTGAACATTGTGGAGAGGTAGTACGGGTTGCAGTCTCAAGGAAAGGACAATATCCAGTTGGATT
TGTCCACTTTGCCAGTCGTACAGAGCTCGACAATGCAATAAAAGAAATGGATGGTGAAACAGTG
AGAGGACCTGACCGAGGGGCAACTTTCAGGATCCAGGTCTCAGTTGCTCGGCCTGTGGTAGAGA
ACGATAAAAAGAGAATTCGTGAAGAAGTGAAAACTAGAAGATCAAACGTATCAACAGACAAGCC
GGACCATTCTTATGGAAGACGTGGACATGATTCATATGATCGTCAAGCAAAAGCTCCAAGGCTA
TATAATGAGGTGTCTGATACGGACCCCTATGAAGCAGCTGTTGTTTCACTACCTTCAGCCGTCA
AGGAACTCCTACTTCGTATTCTACGTCTTAGAATTGGCACTCGATATGATATAGACATTCATTG
CATAAGGAGTCTTAATGAACTTCCTGAAAAGGCTGCAGTTGCTGTCCTTAATCAGTTTTTGATA
TCAGGTGCAGATAAACACAATAAAGGAGACTATTTCGCTTCATTAATTGCTAAGTACCAGGCTG
AGACATTTAGCTCAGCACTAAGATTGCAGGGTTCTACTTATTTGCCAAGAAATCCTGGAATACA
GAACAAGAGATTCCCACATCAAGATTACGAGTACACAGCATCCGGGAGTAGTAGATACAGTTCC
TTAGGTGATTATCCTTCCTCATCTTATGTGGATGATCCCGCATCATCTCAGTCAAGGAATAGAA

FIGURE 6

GGTATGATGAATACAGACCTGATCTTGTAAGATATCCAGATTCAAGATCACGGCAAGAGGAAAT
AGTCCGCATTGAAAGATATCCAGAACCAAGATTTGCACATGAACCAAGACAGGATACTGGAAGG
CATCTCGATCTAGGGTACGTACAAGAACGGAATTCGAATATTGAGAGATCAGCTCAAGTAGCTT
TTTCATCTAGGGAAGGAGGATACTTATCTGCTTCAAGGTACAACACAAACATAGTCCCAGAATT
CAGCTCCAGGTCATCTGCTGAATACTCTACTGCTCGCCAACAAGTAAGGTTTGATCCATTCACA
GGGGAACCATACAAGTTTGATCCCTACACCGGTGAACCCATCAGGCCAGAATCGAACCCACGTC
GCTCAGGAAGCTTATACTGACTTTGATTGATTGAAGCAACAGTTTGGATATGGTAGATTAGATT
TACATCCCTGAACCAAAAGGACCATATACTGCTCTTGCATGTTGTAAACCTAGTGTATTTGATG
TGCCTCAGCATTGTAATGTTAGAAATCCATTTTCATCCATGTCACTGGAAAACTATGGTTGAAA
CAACAGTAATAAGTTCTATCATTTATGATGGCATCTGATGATATGAATTAGGGAAAACTAAGC

seq id no 08: *Oryza sativa* RNA-binding protein (AK059444)

MAKVKTAFVGNLPANVTEEYLRKLFEHCGEVVRVAVSRKGQYPVGFVHFASRTELDNAIKEMDG
ETVRGPDRGATFRIQVSVARPVVENDKKRIREEVKTRRSNVSTDKPDHSYGRRGHDSYDRQAKA
PRLYNEVSDTDPYEAAVVSLPSAVKELLLRILRLRIGTRYDIDIHCIRSLNELPEKAAVAVLNQ
FLISGADKHNKGDYFASLIAKYQAETFSSALRLQGSTYLPRNPGIQNKRFPHQDYEYTASGSSR
YSSLGDYPSSSYVDDPASSQSRNRRYDEYRPDLVRYPDSRSRQEEIVRIERYPEPRFAHEPRQD
TGRHLDLGYVQERNSNIERSAQVAFSSREGGYLSASRYNTNIVPEFSSRSSAEYSTARQQVRFD
PFTGEPYKFDPYTGEPIRPESNPRRSGSLY

seq id no 09: *Oryza sativa* RNA-binding protein DNA (BAC83046.1)

ATGGAACCGACGCGCCGTTGCGTCCCCGGCCATCTCGCCACCGCCGCCGCCGCCGCCGCCGCCT
CGCCGTTCTCCCCGCCGCCGTCGCTGCCGCTGCCGTCCGCGCTCATGCCCCCCAAGAAGCGCCG
CCTCTTCACGCCCGCCCCTCGCCACGCCGCCACCCCGCCACCACCACCACCTCCCCCCACCCCC
GCCGTCGAGCCCACCCTACCAATCCCCCCCGCCTCGACACCGCCGACGCCGCCTCAGCCCTCCG
CCTCCACGGAGCCCTCGACGGCGCCGCCTCCCGCTGTCGACGACGCGGCGGCGAGGTCGTCGTC
GTCGTCGTCGCCGGCGTCGGCGGCGGCGGCGCGGAAGGTTCGGAAAGTGGTTAAGAAGGTCATC
GTCAAGAAGGTCGTCCCCAAGGGCACGTTCGCCGCTCGGAAGGCCGCGGCGGCGGCGGTTGCTG
CTGCTGCGGCGGTCTCCGGAGCAGCAGCATCATCGGAGGCAGGGGGAGAAGCCCCAACCGACGA
GCCAGCAAGTGATCAGGACGGCGGAGTTGGGAATGAGCAAAAATTGGATGAATCCAAACCTGCC
ACGGATTGCAATGCCGTTGCGGTGGTGGAAGAATCGGTGTGTAAGGAGGAGGAGGAGGTGGCCT
TAGTGGTGGGTAAGGGAGTGGAGGAGGAGGAGGCGGGGATGTCGGAGCGGCGGAAGAGGATGAC
CATGGAGGTGTTTGTTGGTGGGCTTCACCGGGACGCCAAGGAGGATGATGTGAGGGCGGTGTTC
GCCAAGGCCGGGGAAATCACCGAGGTCCGGATGATAATGAATCCTCTTGCAGGGAAGAACAAGG
GGTACTGCTTCGTGCGCTACCGCCACGCCGCGCAGGCGAAGAAGGCCATCGCGGAATTCGGCAA
TGTGAAGATTTGTGGGAAGCTCTGTCGAGCTGCAGTTCCAGTTGGGAATGACAGAATTTTTCTT
GGAAACATCAACAAGAAATGGAAAAAAGAAGATGTCATCAAGCAGCTAAAGAAAATTGGAATTG
AGAACATTGATTCTGTAACACTTAAGTCTGATTCAAATAATCCAGTCTGTAATCGTGGTTTTGC
ATTTCTTGAACTGGAAACTAGTAGAGATGCACGGATGGCATACAAAAAGCTTTCACAGAAAAAT
GCTTTTGGCAAAGGCCTGAATATAAGAGTTGCATGGGCTGAACCATTGAATGATCCAGATGAGA
AAGATATGCAGGTTAAATCGATTTTTGTGGATGGGATACCAACGTCCTGGGATCATGCTCAGCT
AAAAGAAATCTTCAAGAAACATGGGAAGATTGAAAGTGTGGTTCTGTCACGCGATATGCCGTCA
GCTAAAAGGAGGGACTTTGCCTTTATTAATTACATTACTCGTGAGGCTGCAATCTCGTGTCTTG
AATCTTTTGACAAGGAAGAGTTCAGTAAGAACGGCTCAAAGGTGAATATTAAAGTTTCATTGGC
TAAACCTGCCCAACAGAGCAAGCAGACCAAGGAAGACCATAAATCTAGTATTACTGGGGAAGGC

FIGURE 6

AAAATGAAGACTTCTAAAATAAGATACCCTGTTCAAGATTATACCCACATTTATTCTGGAGAGA
AGCGTCCCTTTTCAACACTGGGTGATCCTTATTATCCATTGAGAGGTCATTCTTGTCGTCGTCA
TGAGGGTAGCACCTATACTACAGCAGCATCAAGCTATGGTGCGCTGCCCCCTGCTACTGCTGAA
TCTTCTCTGCCACATTATCATGACAGCAATAGATATCCTCCACACCTAGGTGAGGCAATCAAGT
TCTCGCCAACCAGCGCAGTCCTATCGAAGCAGGCATGGCAAAAAATGTAA

seq id no 10: *Oryza sativa* RNA-binding protein (BAC83046.1)

MEPTRRCVPGHLATAAAAAAASPFSPPPSLPLPSALMPPKKRRLFTPAPRHAATPPPPPPPTP
AVEPTLPIPPASTPPTPPQPSASTEPSTAPPPAVDDAAARSSSSSSPASAAAARKVRKVVKKVI
VKKVVPKGTFAARKAAAAAVAAAAAVSGAAASSEAGGEAPTDEPASDQDGGVGNEQKLDESKPA
TDCNAVAVVEESVCKEEEEVALVVGKGVEEEEAGMSERRKRMTMEVFVGGLHRDAKEDDVRAVF
AKAGEITEVRMIMNPLAGKNKGYCFVRYRHAAQAKKAIAEFGNVKICGKLCRAAVPVGNDRIFL
GNINKKWKKEDVIKQLKKIGIENIDSVTLKSDSNNPVCNRGFAFLELETSRDARMAYKKLSQKN
AFGKGLNIRVAWAEPLNDPDEKDMQVKSIFVDGIPTSWDHAQLKEIFKKHGKIESVVLSRDMPS
AKRRDFAFINYITREAAISCLESFDKEEFSKNGSKVNIKVSLAKPAQQSKQTKEDHKSSITGEG
KMKTSKIRYPVQDYTHIYSGEKRPFSTLGDPYYPLRGHSCRRHEGSTYTTAASSYGALPPATAE
SSLPHYHDSNRYPPHLGEAIKFSPTSAVLSKQAWQKM

seq id no 11: *Oryza sativa* prolamin promoter

CTTCTACATCGGCTTAGGTGTAGCAACACGACTTTATTATTATTATTATTATTATTATTATTAT
TTTACAAAAATATAAAATAGATCAGTCCCTCACCACAAGTAGAGCAAGTTGGTGAGTTATTGTA
AAGTTCTACAAAGCTAATTTAAAAGTTATTGCATTAACTTATTTCATATTACAAACAAGAGTGT
CAATGGAACAATGAAAACCATATGACATACTATAATTTTGTTTTTATTATTGAAATTATATAAT
TCAAAGAGAATAAATCCACATAGCCGTAAAGTTCTACATGTGGTGCATTACCAAAATATATATA
GCTTACAAAACATGACAAGCTTAGTTTGAAAAATTGCAATCCTTATCACATTGACACATAAAGT
GAGTGATGAGTCATAATATTATTTTCTTTGCTACCCATCATGTATATATGATAGCCACAAAGTT
ACTTTGATGATGATATCAAAGAACATTTTTAGGTGCACCTAACAGAATATCCAAATAATATGAC
TCACTTAGATCATAATAGAGCATCAAGTAAAACTAACACTCTAAAGCAACCGATGGGAAAGCAT
CTATAAATAGACAAGCACAATGAAAATCCTCATCATCCTTCACCACAATTCAAATATTATAGTT
GAAGCATAGTAGTA

seq id no 12: Artificial sequence - motif I (consensus sequence)

PYEAAVVALPVVVKERLVRILRLGIATRYD

seq id no 13: Artificial sequence - motif II (consensus sequence)

RFDPFTGEPYKFDP

seq id no 14: At1g58470 (contig f9k23 - coordinates: 4229-5089, 5173-5394)

aagatttgggcttacaatctttatcacaaaggctttttaaagcccattagttacattcatcat
tatctctcgacattaaaaaaaaaaagttaaactgaagaagctaaaaagagttttaacttttaa
ctctcttcgtcttctccctcgtgccgtgtcaaatcaatctactgttctctctcctatctggtaa
acttttcctcttcgccatgaaatttttttcttgctagggttttagtttctacagttcgcttccc
aaaaattaggggttttgtcacaatttctcaatttcttgttccatttttcttcttttctccataa

FIGURE 6

tcattgcttaatttagaatcccaaattttacaaattagggtttttgtttaattttaggggtttt
tgattttcaactgttaatagtgttctcgatgtcataattctgatttttttttattatctattccg
aaattagggcaaaaatctcagacaaacctgcaaaattagggtatttgaggatATGGATTATGAT
CGGTACAAGTTATTTGTTGGTGGTATTGCGAAAGAGACAAGTGAAGAAGCTCTGAAGCAGTATT
TTAGCAGATATGGAGCTGTGTTGGAAGCTGTTGTAGCTAAAGAGAAAGTCACTGGAAAACCTAG
AGGTTTTGGGTTTGTTCGCTTTGCTAATGATTGTGATGTTGTTAAAGCTCTTAGAGACACTCAC
TTCATTCTCGGTAAACCCgtaagtgttaccgccttttttatgcttgtgtcaattgggttttgtgt
atactctgtggattgattatgtgtgtgtttgtattagGTTGATGTGAGAAAGGCGATTAGGAAA
CATGAACTATACCAACAGCCGTTTAGCATGCAGTTTTTGGAGAGAAAAGTGCAACAGATGAATG
GTGGTTTGCGTGAGATGTCGAGTAATGGTGTGACCAGTAGGACTAAGAAGATATTTGTTGGGGG
TTTGTCGTCTAACACGACTGAGGAAGAGTTTAAGAGTTACTTTGAGAGGTTTGGTAGGACTACT
GATGTAGTTGTGATGCATGACGGTGTGACTAACAGGCCAAGGGGTTTTGGGTTTGTTACTTATG
ATTCGGAGGACTCTGTTGAGGTTGTTATGCAGAGTAATTTCCATGAGTTGAGTGATAAACGCGT
GGAAGTGAAACGGGCAATACCTAAAGAAGGAATCCAGAGCAATAACGGTAATGCTGTTAATATT
CCTCCTTCCTACAGCAGCTTTCAAGCAACACCTTATGTCCCTGAGCAAAACGGATATGGGATGG
TTTTACAGTTTCCTCCTCCTGTCTTTGGTTATCATCACAATGTCCAAGCCGTTCAATATCCTTA
TGGTTACCAATTCACAGCACAAGTGGCTAACGTTTCATGGAACAATCCGATTATGCAACCCACC
GGTTTTTACTGTGCTCCTCCTCATCCTACTCCTCCTCCCACCAACAATCTTGGTTATATCCAAT
ACATGAACGGGTTTGATCTTTCGGGTACGAACATTTCCGGGTACAATCCTCTAGCATGGCCTGT
AACGGGGGATGCAGCTGGTGCGCTAATACATCAGTTTGTAGATTTGAAGCTTGATGTCCACAGT
CAAGCCCATCAGAGAATGAATGGAGGTAACATGGGAATACCATTGCAGAATGGTACATATATAT
GAcagttgcagaatgataaatgcaaataggctcacaagggtagtgaaattctttggactctttt
aaatggtttttaggttcctcatctttcttcattaactctttggtaaatgtgttgggttggttt
ggttaccttgtatattgtttaggtatttgattttaaccccaagacttatgtatcatatattact
gcatttgtaatatatcacactcatttagttcattttgttgcttttatggttttgttgattttgt
ggtttcgttgattaaattggcaatgatgttttaaattcatcaaggaaaacaaagaaatagattg
tcgattaaacagtagaaaaaggaaatagttttgtagaaataggaactgaatctggaaatctcta
agaataccatattgtagaaagaaaataaatctgagacgggagaaactatcgagcatccttgagc
tttaagttggagaaaccgggtaagcgtttgtgggattttgttgtaagattgaac

seq id no 15: At1g58470 protein sequence

MDYDRYKLFVGGIAKETSEEALKQYFSRYGAVLEAVVAKEKVTGKPRGFGFVRFANDCDVVKAL
RDTHFILGKPVDVRKAIRKHELYQQPFSMQFLERKVQQMNGGLREMSSNGVTSRTKKIFVGGLS
SNTTEEEFKSYFERFGRTTDVVVMHDGVTNRPRGFGFVTYDSEDSVEVVMQSNFHELSDKRVEV
KRAIPKEGIQSNNGNAVNIPPSYSSFQATPYVPEQNGYGMVLQFPPPVFGYHHNVQAVQYPYGY
QFTAQVANVSWNNPIMQPTGFYCAPPHPTPPPTNNLGYIQYMNGFDLSGTNISGYNPLAWPVTG
DAAGALIHQFVDLKLDVHSQAHQRMNGGNMGIPLQNGTYI

seq id no 16: At4g26650 (contig wt_d_33 - coordinates: 74595-74606, 74780-75016, 75420-76524, 76622-76635)

cttcattgagagagagatatagagagagaaaagagagagaggccatattttgataagagaagaa
gaacccttatagagaaagagaaagagagagacagagagagtggatggatgtcttatagaatgaa
caaaacatcctctgtttctcttgtccttgtccctttttccagatcttaaggttttccacatttt
atcatctgggtcctctccttaatggtgaattctccatctttacaagtttgatgttttttgttcat
caaatctggcgtttttttttctcttctaatatatattgtctctgctcattttccgtttctcttc

FIGURE 6

ccattgattgttctgtttcatttctgtttttttttttcaatagttttgattggatgctttgat
gatccattgtcagatttgaagacactcaattcctatttgatcggggactagaatttggattctg
tttcagacaaaagtagatttccctgtctctttcccgtttgattttcaataagATGAATCCGGAG
gtaaaacattgaacaattcttcataaatctcagaactttgagctttttgaatcttaaaacacg
atcgaagtaaaaaatcgaattgttagatgaaatgggcaatcgtcattttcgcaaatctgatccg
tatttgtgagatcggattcattggatcgactttggggttttgcagGAGCAAAAGATGGAATCTG
CATCGGATCTGGGCAAGCTCTTCATTGGCGGGATTTCATGGGACACAGATGAAGAACGACTGCA
AGAGTATTTTGGCAAGTATGGAGATTTGGTTGAAGCTGTGATCATGAGAGACCGTACTACCGGA
CGTGCCCGTGGCTTTGGGTTTATCGTTTTTGCAGATCCTTCTGTTGCCGAGAGAGTCATCATGG
ACAAACACATCATTGATGGCCGCACGgttagtattcttggatccattgcttgacaattcatcta
attatcagtcttgagtaatcgagtgttctaaagtctcgatctttctgtaatgattctgtcttag
aggtcttattggtctcgctgctcgttaatgagcaacggattgttctataatctcgatctttctg
tattcatgctctcttagagatctgtttggtgtcatccattaatgagttttaagcagcaacgttt
agatctttctgtaatcatgctcttttcgaaatcttctgttgtcattagcttctggatttgctgt
tactgttataacttgtgagaatgtgttgttgctttgtgttgaagtggcaatgttagtgttagat
caatgagaaaagaatgaaagatctttttttatttctttgttgcagGTCGAGGCGAAGAAAGCTG
TCCCGCGGGATGATCAGCAAGTGCTAAAACGACACGCCAGTCCAATGCACCTTATCTCACCTAG
CCATGGTGGTAATGGTGGTGGAGCACGGACAAAGAAGATCTTTGTTGGAGGTTTACCGTCTAGC
ATTACTGAGGCCGAGTTCAAGAACTACTTTGATCAGTTTGGTACAATTGCTGATGTTGTGGTAA
TGTATGATCATAATACACAGAGGCCAAGAGGCTTTGGCTTCATCACTTTTGATTCCGAAGAGTC
TGTTGATATGGTTCTCCACAAGACCTTTCATGAGCTAAACGGAAAAATGGTTGAGGTTAAAAGA
GCAGTGCCAAAGGAGCTCTCCTCGACTACTCCTAACCGAAGCCCACTTATTGGGTATGGTAACA
ACTATGGAGTAGTCCCTAATAGGTCTTCTGCTAATAGCTACTTCAATAGTTTTCCTCCTGGTTA
TAATAATAATAATCTAGGCTCTGCTGGCCGGTTTAGTCCTATTGGTAGCGGTAGAAATGCTTTC
TCTAGCTTCGGGCTCGGATTGAATCAAGAACTGAATTTGAATTCAAACTTTGATGGAAACACTC
TTGGGTATAGCCGGATCCCTGGCAACCAATACTTCAACAGTGCTTCACCAAACCGTTACAACTC
TCCAATTGGGTACAACAGAGGAGACTCTGCTTACAACCCGAGCAACAGAGACTTGTGGGGAAAC
AGAAGCGATTCCTCTGGTCCAGGTTGGAACTTGGGAGTTTCGGTTGGTAACAACAGAGGAAACT
GGGGACTTTCTTCTGTGGTGAGCGATAACAATGGCTATGGAAGAAGCTATGGGGCTGGTTCTGG
ACTTTCGGGGTTATCATTCGCGGGTAATACAAACGGTTTTGATGGCTCTATAGGGGAATTGTAT
AGAGGCAGCTCAGTTTATAGCGACTCAACATGGCAGCAGTCAATGCCTCATCATCAGTCTTCTA
ATGAGTTAGACGGCTTGTCTCGCTCTTATGGCTTTGGTATTGACAATGTAGGCTCAGACCCATC
AGCCAATGCCTCAGAAGGATACTCCGGAAACTACAATGTCGGAAATAGACAAACACATAGAGgt
acactcatcgatgtcaaacttttttccttttgcatctcatctgctacatttatttttgcctgtt
gaaaagtaattagattgattaacgttttcagGTATTGAAGCATAGaaagaaatcgacgaagaga
agtgagaattgtagatcaagaagaacagccatttccgttgcagagtttgaagagttgttatttc
gatatcaagtagagaaagaaaccaactttcttcatcacagtgagtttcttgttttgttttttttc
gtcgttagcatcacaaacacaaaaaagagaagtttatttttacttaaaaattcttacataaga
taagatcagattggtagctgcaaagatacaacatggatgataaaaaaagatttggtttcgtctc
catagcaataaccagagatcgttgattctcgatcactattctttaggtttctctccttcttctt
ccatgatttcttgatgttgtgtgctctgtttgtaactctaattgttaaaatttttttatgttaca
gatttttttttttctttttggtttttaaactttggattcgaattgttcatgggaacttttggattt
ttctattagcgtgagagaaaacacattgtgcaa

FIGURE 6

seq id no 17: At4g26650 protein sequence

MNPEEQKMESASDLGKLFIGGISWDTDEERLQEYFGKYGDLVEAVIMRDRTTGRARGFGFIVFA
DPSVAERVIMDKHIIDGRTVEAKKAVPRDDQQVLKRHASPMHLISPSHGGNGGGARTKKIFVGG
LPSSITEAEFKNYFDQFGTIADVVVMYDHNTQRPRGFGFITFDSEESVDMVLHKTFHELNGKMV
EVKRAVPKELSSTTPNRSPLIGYGNNYGVVPNRSSANSYFNSFPPGYNNNNLGSAGRFSPIGSG
RNAFSSFGLGLNQELNLNSNFDGNTLGYSRIPGNQYFNSASPNRYNSPIGYNRGDSAYNPSNRD
LWGNRSDSSGPGWNLGVSVGNNRGNWGLSSVVSDNNGYGRSYGAGSGLSGLSFAGNTNGFDGSI
GELYRGSSVYSDSTWQQSMPHHQSSNELDGLSRSYGFGIDNVGSDPSANASEGYSGNYNVGNRQ
THRGIEA

seq id no 18: At5g55550 (contig mte17 - coordinates: 77934-79094, 79236-79457)

atatgtgagactaactattgttctctgtctctttttttctttttaattatcaaagaaagaaact
ctttcttaatggaaaccatttacagataaaaaaaacattaaaaggaaaggtttttaataaagcc
tttgagagagaagatgtttattataggatgaacaaaaacatcctctgtttctctcttttcatat
ttttctccacatttcctcatctgggtcatctccaaaaatggtgctttttttttaataattcttca
cgtttctgggtttttggttttgtgatttgatgatgctttttttttgttttttttcagatttgatg
ataacccaaattcgcaatttgattaggacaacaacaacaacttttatttatctgattccgtcttt
gattttcagacaagaaaagtatgttgtttctaagtcttttgatttttttcaatttcatctcctt
actcgatttttttttttttgggtttctctgaattggagcagaaaaaaaaaagATGGAATCGGAT
CTGGGGAAGCTCTTCATTGGTGGGATTTCGTGGGATACAGACGAAGAAAGGTTAAGAGACTACT
TTAGCAACTATGGTGATGTTGTTGAAGCTGTGATCATGAGAGATCGTGCCACAGGTCGTGCACG
TGGCTTCGGCTTCATTGTCTTTGCAGACCCCTGTGTCTCAGAGAGAGTGATCATGGATAAACAC
ATCATCGATGGCCGCACGgtttgtgatttcaatcatttctcaatctttcagcagaacaaacaaa
gttcagatcttattgcaacttcctcaatttgcgtttttgaatcatctctcaatctttgtttctc
aaagtgtaaagatcaaatttatgttttgcagGTTGAGGCGAAGAAGGCTGTGCCTCGAGATGAT
CAGCAGGTGCTAAAGCGACACGCTAGTCCTATCCACCTTATGTCACCTGTCCATGGTGGTGGTG
GAAGGACAAAGAAGATCTTCGTTGGAGGTTTACCGTCTAGCATTACCGAGGAGGAGTTCAAGAA
CTACTTTGATCAGTTTGGTACTATTGCTGATGTTGTTGTAATGTATGATCATAACACGCAGAGG
CCAAGAGGTTTTGGCTTCATCACATTTGATTCAGATGATGCTGTTGATAGAGTTCTTCACAAGA
CCTTCCATGAGCTCAATGGGAAACTAGTTGAGGTCAAAAGAGCTGTACCTAAGGAGATTTCCCC
TGTTTCTAATATCCGAAGCCCGCTTGCTAGCGGTGTTAACTATGGAGGCGGGTCTAATAGGATG
CCTGCTAATAGCTACTTTAACAACTTTGCTCCTGGTCCTGGTTTTTATAACAGTCTAGGTCCTG
TTGGTCGTCGGTTTAGTCCTGTTATTGGTAGTGGTAGAAATGCGGTTTCTGCTTTTGGCCTCGG
TTTGAATCATGACTTGAGTTTGAATTTGAATCCAAGCTGCGATGGGACAAGTTCTACGTTTGGT
TATAACCGTATTCCAAGCAACCCTTACTTCAACGGTGCTTCCCCGAACCGTTACACCTCTCCAA
TCGGGCACAATAGAACTGAGTCTCCTTACAATTCGAACAATAGAGACTTATGGGGAAACAGAAC
CGACACTGCAGGTCCCGGTTGGAACTTGAATGTCTCGAATGGAAACAACAGAGGAAATTGGGGA
CTTCCTTCTTCTTCTGCTGTTAGTAATGATAACAATGGCTTTGGAAGGAACTATGGGACAAGTT
CTGGACTTTCCTCGTCCCCATTTAATGGTTTTGAAGGTTCTATAGGGGAACTGTACAGAGGCGG
CTCAGTCTACAGCGACTCAACGTGGCAGCAACAGCAGCTACCATCTCAGTCTTCTCACGAGCTA
GACAATTTGTCTCGCGCTTACGGTTATGATATTGACAATGTAGGTTCAGACCCATCTGCAAATG
ACCCAGAAACTTACAATGGAAGCTACAATGTTGGAAATAGACAAACTAATAGAGGTAACAAAAA
AATTCATCTCAATAAAACTTGTAACTTGGATACATTTTGAtcgcaatcgaaatgttctgatctg
tgttttatttacttgttgaggtattgctgcataggttatcaaaaaccaagaaaacaaaaaaaaa
agttgagagatttgtagattgaaagcaaccaaatttcagttgcagagtttgaacaggttctcat

FIGURE 6

gacaaagaaaccaactttgttgatcacagtgccaaagattatggtttgctttctcttttgttag
accaaaaaaaaaaaaaaaaaagagaaaaacaaagaaccgtttttgtttttcttcttcttacata
aagatcagatcgtagcagccagacaaccaaagatactacaaggtggatttagatttgcttctca
aaaaagtttttttttttctttcatagaataaccaaacaaagatcgtagaattttcgatcaaaga
ttcttcagagttctgtgctctgttttgtaattgtacttttttttttcttgtttacaaaatgaatt
gttcatgaaaactttgttttcttaaaaa

seq id no 19: At5g55550 protein sequence

MESDLGKLFIGGISWDTDEERLRDYFSNYGDVVEAVIMRDRATGRARGFGFIVFADPCVSERVI
MDKHIIDGRTVEAKKAVPRDDQQVLKRHASPIHLMSPVHGGGGRTKKIFVGGLPSSITEEEFKN
YFDQFGTIADVVVMYDHNTQRPRGFGFITFDSDDAVDRVLHKTFHELNGKLVEVKRAVPKEISP
VSNIRSPLASGVNYGGGSNRMPANSYFNNFAPGPGFYNSLGPVGRRFSPVIGSGRNAVSAFGLG
LNHDLSLNLNPSCDGTSSTFGYNRIPSNPYFNGASPNRYTSPIGHNRTESPYNSNNRDLWGNRT
DTAGPGWNLNVSNGNNRGNWGLPSSSAVSNDNNGFGRNYGTSSGLSSSPFNGFEGSIGELYRGG
SVYSDSTWQQQQLPSQSSHELDNLSRAYGYDIDNVGSDPSANDPETYNGSYNVGNRQTNRGNKK
IHLNKTCNLDTF

seq id no 20: At4g14300 (contig fca_all - coordinates: 189072-189220, 189554-189611, 189665-190678)

ctgtaatgtggagtttggaattttcgacaacaaagtgcacatctggcacagagattgtcacagc
acgaaagatttttttgtcgttcttgtaggatttgctggcacgtgtggaatagaaaacacacgag
tgaaaccatcgtcggtctttgtagcccattatttatacttctattgggctggacttaagcccat
aagtaagcatctctgttacaagaaaacgggaaacagatctgaaccgttaataatattagaaagg
atctagaccgttgatttatttatctgctgacagattcgtaccttcgcgaatatcaataccaaac
caatagaaatattcgttcgctgtcttcttcctcttcctcctctcaaatcggctacagccattgg
aaaagctaaagccttttcgtaatttctggaagtttctgcagtcggttttcacggtttcgtagat
tgaggtggatttgtgattctgggtcagaagtaagatagtggaatataaattcATGGATTCGGAT
CAAGGAAAGCTTTTTGTCGGTGGTATTTCATGGGAAACTGATGAAGATAAGCTGAGAGAACATT
TCACCAACTATGGAGAGGTTTCTCAGGCTATTGTGATGAGAGACAAGCTCACAGGTCGACCTAG
GGGTTTTGGgttcgttatcttctcggatccttctgttctcgatagggttcttcaagagaaacac
agcattgataccagagaggttattattgttctcttatagctccatttctctaattgtgttaaag
ttttatccttttgcgttttgctgtgttgattgagaacgagagtaaatatagaattttgtttgg
ttggcaaattcgccttagtgtttcttagattctaggattggttttaacttgtataagaggtatt
atagggtactcgatatatgttaatcgtacactctatgaagtgattgagtatagtattagaaaag
agagcttggtttggtttattagGATAAGGAAAAACAGATGTATATATTTTCTGTTGCGTTATGT
TCTCGATTTGGGTAAAgtatgattcttggaagtttattatgagctttattgattttggttaatg
tttagGTTGATGTGAAGAGAGCCATGTCAAGAGAGGAGCAGCAAGTCTCTGGAAGAACTGGGAA
TCTTAATACATCTAGAAGTTCTGGAGGTGATGCTTACAATAAAACCAAGAAGATCTTTGTTGGA
GGCTTGCCACCTACTTTGACTGATGAAGAGTTTCGCCAGTACTTTGAAGTTTATGGCCCTGTGA
CTGATGTTGCAATCATGTATGACCAGGCTACCAACCGTCCTCGTGGGTTTGGATTTGTTTCCTT
CGACTCTGAAGATGCGGTAGACAGTGTTTTGCACAAGACTTTCCATGATTTGAGCGGTAAACAA
GTTGAAGTAAAGCGTGCTCTTCCTAAAGATGCCAATCCTGGAGGTGGTGGACGATCAATGGGTG
GTGGTGGCTCTGGTGGTTACCAGGGTTATGGTGGCAATGAAAGCAGTTATGATGGACGTATGGA
TTCCAATAGGTTTTTGCAGCATCAAAGTGTTGGAAATGGTTTACCATCTTATGGTTCTTCTGGT
TATGGCGCTGGCTATGGAAATGGTAGTAATGGTGCCGGGTATGGTGCCTATGGAGGTTACACTG
GTTCTGCTGGAGGTTATGGCGCTGGTGCTACTGCTGGATATGGAGCAACGAACATTCCAGGTGC

FIGURE 6

TGGCTATGGAAGTAGTACTGGAGTTGCTCCGAGAAACTCATGGGACACTCCAGCTTCTAGTGGT
TATGGGAACCCAGGCTATGGGAGTGGTGCTGCTCATAGTGGATATGGAGTTCCTGGTGCAGCTC
CTCCTACGCAGTCACCATCTGGCTATAGTAACCAAGGCTACGGTTATGGAGGGTACAGTGGAAG
TGATTCTGGTTATGGAAATCAAGCTGCATATGGTGTGGTTGGAGGGCGTCCTAGTGGTGGCGGT
TCAAACAACCCTGGTAGTGGTGGCTACATGGGAGGTGGTTATGGTGATGGATCTTGGCGATCTG
ACCCGTCACAAGGTTATGGTGGTGGGTACAATGATGGTCAGGGTCGACAAGGCCAGTAGtgact
gtgtaaggggattatgaccgccctggtttctggatccttgtcaagaagaatttagctcaaatca
aaggttccacaacttcctaacgggttggactgcttgaatctctttataagcatgtgctatctat
tacaataagtcacttctattaagttatttttcggttgagtgtacttttgagttttggcagagtt
attataactacaggctttgctgttttcgtattatgtttgtcttcctagtattcttgccggattg
tttgttttgattgtgttatttttgttttggccctgatggatataacttaagcagggaataatgct
tcagggtacttgttaagaaagcagatggtgagagcagaactcgatggaggtgagagtcaaattg
ctgaatgtatggtttgagtagaaagtagaggtagttggtaacgttagtggtaccattaagaaga
aggtgtagaaaatagtgagaggtagctttgagaaaaaggcataatca

seq id no 21: At4g14300 protein sequence

MDSDQGKLFVGGISWETDEDKLREHFTNYGEVSQAIVMRDKLTGRPRGFGIRKNRCIYFLLRYV
LDLGKVDVKRAMSREEQQVSGRTGNLNTSRSSGGDAYNKTKKIFVGGLPPTLTDEEFRQYFEVY
GPVTDVAIMYDQATNRPRGFGFVSFDSEDAVDSVLHKTFHDLSGKQVEVKRALPKDANPGGGGR
SMGGGGSGGYQGYGGNESSYDGRMDSNRFLQHQSVGNGLPSYGSSGYGAGYGNGSNGAGYGAYG
GYTGSAGGYGAGATAGYGATNIPGAGYGSSTGVAPRNSWDTPASSGYGNPGYGSGAAHSGYGVP
GAAPPTQSPSGYSNQGYGYGGYSGSDSGYGNQAAYGVVGGRPSGGGSNNPGSGGYMGGGYGDGS
WRSDPSQGYGGGYNDGQGRQGQ

seq id no 22: At3g07810 (contig f17a17 - coordinates: 43905-44126, 44339-45587, 46069-46082)

ttgaaattgggttaaatcggtttgaatcggattgaacaaaaactgtattaataataattcttcc
tctacttttctctctgattgattccaatcttctttcattttcttcttcttcttctggggaa
ggggcaggttaaaattatgccatctattcaaatcgtgcctatcctcagatcttaactcttttct
ctacgagattcggcatctgggttttattcttcttggtgggtttttttttattcttcttcttctg
atctcagatttcccctgattggtttttttttttgctaaatccgttttatgttttcccgatcaaa
ctctcctggcagattctcggatctgttgttttctagattcaatctgaatttgattttacgtttt
tgtctttgtaaagatgtttccttttgatcagattttgataatccattgacatctctgattcaag
caaaagctaattaactttgatccgattcctttgtgtgtgtgtgcagagcaaaATGCAATCGGAT
AATGGAAAGCTTTTCATCGGTGGGATATCTTGGGACACCAATGAGGAACGTCTCAAGGAGTATT
TCAGCAGTTTTGGAGAAGTGATCGAAGCTGTCATCTTGAAAGATCGTACCACTGGTCGTGCTCG
TGGTTTCGGTTTTGTTGTTTTTGCTGATCCTGCTGTTGCTGAGATTGTTATCACCGAAAAACAT
AATATTGATGGCAGATTGgtatgttcactgttctctgcctttcgttttgtacaatgtaacttg
ttttcgaagcttccttatgcaatcaagccttcaagagttacagtttgttctcatttggttccga
ttaatcatttttgtgctttgattggatttttgagaagaaatgagtgatctttagttatatgagc
ttagttttcatttttcaagttgtttgatcttccgcagGTTGAAGCCAAGAAAGCTGTTCCCAG
AGATGACCAAAACATGGTAAATAGAAGCAACAGCAGTAGCATCCAAGGTTCTCCCGGTGGTCCA
GGTCGCACAAGGAAGATATTTGTTGGAGGATTACCTTCTTCGGTTACAGAGAGTGATTTCAAGA
CGTATTTTGAGCAGTTTGGTACAACTACGGATGTGGTTGTCATGTATGATCACAACACACAAAG
GCCTAGAGGTTTCGGGTTTATAACCTACGATTCCGAGGAGGCGGTTGAAAAGGTATTGCTCAAG
ACATTCCATGAACTAAATGGTAAAATGGTTGAGGTTAAGCGAGCTGTTCCAAAGGAGTTATCTC

FIGURE 6

CAGGTCCAAGTCGCAGTCCTCTTGGTGCAGGTTACAGCTATGGAGTTAATAGGGTCAATAACCT
CCTTAATGGGTATGCTCAAGGGTTTAATCCCGCTGCAGTTGGAGGCTACGGACTTAGGATGGAT
GGTCGGTTCAGTCCGGTTGGTGCTGGAAGAAGCGGGTTTGCAAATTACAGTTCTGGATACGGGA
TGAATGTGAACTTTGATCAGGGATTGCCCACAGGGTTCACGGGAGGTACAAATTACAATGGAAA
TGTTGACTATGGCCGAGGAATGAGCCCGTACTACATTGGTAACACAAACAGGTTTGGTCCTGCG
GTTGGCTATGAAGGGGGCAACGGAGGAGGAAACTCATCCTTCTTCAGTTCGGTTACACGGAACT
TATGGGGAAACAATGGTGGTCTTAACTATAACAACAATAATACAAACTCAAACTCCAATACATA
TATGGGAGGATCATCAAGTGGGAACAACACACTTAGTGGTCCATTTGGAAATTCAGGAGTCAAT
TGGGGTGCTCCTGGAGGAGGAAACAATGCTGTGAGTAACGAGAATGTGAAGTTTGGTTATGGAG
GAAACGGTGAATCTGGTTTTGGGTTGGGAACAGGTGGTTATGCAGCAAGAAACCCAGGGGCTAA
CAAGGCAGCACCATCCTCTTCATTCTCTTCTGCCTCAGCAACCAACAACACGGGTTATGATACA
GCAGGACTTGCAGAGTTTTACGGGAATGGTGCAGTTTATAGTGACCCTACATGGAGATCACCAA
CTCCTGAGACAGAAGGGCCTGCTCCTTTTAGCTATGGGATTGGAGGAGGGGTTCCTTCTTCAGA
TGTTTCAGCTAGAAGTTCATCTCCAGGTTATGTTGGCAGTTACAGTGTGAACAAGAGACAACCA
AACAGAGgtaattgagttcagagtaattttctgctttaacatgtgattctatgaaaagcaaagg
actcttgagaaaaagaatttagaaagcctagatagtttccaaatttttgattatcctcgtcttc
tttctggaatatacaaaccatggtttagggtcttgcactaatggtgatctagaacaccttcgta
tcactagtgaattggcttttcctcagaaacacgaatatacttgcatgcagaaacagtagccatt
ctgcatctttattgtttttttagttcatcagagattatttagaggaaagtttctttccgtgcttt
agatataagctcatggaactagaaaactagttgaatcttttatgttgctcacaccagtgtctat
gggaagtctaagaaacttgtgatgaagaaactcaattgcatgactggtttcttatcgctcttct
cttctctgaattatatttccctttttcggttttgttgcagGAATTGCTACTTAGtacaatcgtt
tttgttttaccacgatattgtaggcgagccatcacggtgaacgatctgtgtcttttggcgaatc
ttttagattatcttctttttcccttcatacaaagccagtgaggacgaaacttgatcatatcatca
cctagagctaaccagagaatcccgcagacttttctgtcatggtttggttttctaaattcattgt
tcctcctaggctttttttctgctttctttttttttctatttttgttttctttttcttctttcaatg
agggacagaagaaactgtatcagtctccggcgaggcggtaatacataaggagagttcaaaacaa
aaacccaaaaaaaaaaaaaaaaaagatgatccttcttcctcagttttcttcttcattgtcatgta
atggttcttcttcttttcttcttcttgggggttatggttaaggtttgtgttttgaggcagattg
tactagagttttttttcatgtttcttttgttttgtcgttttt

seq id no 23: At3g07810 protein sequence

MQSDNGKLFIGGISWDTNEERLKEYFSSFGEVIEAVILKDRTTGRARGFGFVVFADPAVAEIVI
TEKHNIDGRLVEAKKAVPRDDQNMVNRSNSSSIQGSPGGPGRTRKIFVGGLPSSVTESDFKTYF
EQFGTTTDVVVMYDHNTQRPRGFGFITYDSEEAVEKVLLKTFHELNGKMVEVKRAVPKELSPGP
SRSPLGAGYSYGVNRVNNLLNGYAQGFNPAAVGGYGLRMDGRFSPVGAGRSGFANYSSGYGMNV
NFDQGLPTGFTGGTNYNGNVDYGRGMSPYYIGNTNRFGPAVGYEGGNGGGNSSFFSSVTRNLWG
NNGGLNYNNNNTNSNSNTYMGGSSSGNNTLSGPFGNSGVNWGAPGGGNNAVSNENVKFGYGGNG
ESGFGLGTGGYAARNPGANKAAPSSSFSSASATNNTGYDTAGLAEFYGNGAVYSDPTWRSPTPE
TEGPAPFSYGIGGGVPSSDVSARSSSPGYVGSYSVNKRQPNRGIAT

seq id no 24: At2g33410 (contig f4p9 - coordinates: 71729-71950, 72087-73079)

atgatctaacatttttctcaaataataaggtcattgatccttatataacatggaatcactata
acatttataacctacattcttgctcatatatctctctccttttttttccaacatattaacgact
aataataaaatttatcaaccattttaaatctctaaatggaacttattattacatgactaaaaaa

FIGURE 6

taaaaataaataaataaataaacgaagctgatatggaaaagtcttctctttctttttttttttt
tggtaagtcgatctctctttcactcactttaacccaattggccgctattttccaaagtctgttt
atttttttaatctctctctcttctctctcacccaatttcacaaacccgaaaccctaattttctc
gggacactgaaatttttacagcttctttcctcttcttcaccggggagatttgtcggtactaaat
ctagggtttttgggtatcaccggagggttgaagagagagaaaaaaactcacaATGGAATCAGAT
CAGGGAAAGCTATTTATCGGCGGGATTTCATGGGATACCGACGAGAATCTTCTGAGAGAGTACT
TCAGCAATTTCGGCGAGGTTTTGCAGGTCACTGTTATGCGAGAGAAAGCTACTGGTCGTCCTAG
AGGATTCGGATTCGTCGCATTCTCGGATCCTGCTGTTATTGATAGGGTTCTTCAGGACAAGCAC
CATATTGATAATAGAGATgtaagcaaaaatcttgtttctcaaatgggtctttctaaattttgaa
tctttatagtaaaaattgatactttgaatcttgttgttgtcgaggtttgattttcatctttgat
ggatttaagttgtgttaatttcttagGTTGATGTGAAGAGAGCAATGTCTAGAGAGGAGCAGAG
TCCTGCTGGGAGATCAGGGACTTTTAATGCTTCTAGGAATTTTGATAGTGGAGCTAACGTGAGG
ACTAAGAAGATATTCGTGGGAGGTTTGCCTCCTGCATTAACATCAGATGAATTTCGGGCTTACT
TTGAGACTTATGGTCCTGTGAGTGATGCAGTCATTATGATTGATCAGACTACACAGCGTCCTCG
AGGATTTGGGTTTGTTTCTTTTGATTCTGAAGATTCGGTTGACCTTGTTTTACATAAGACTTTC
CACGATTTGAATGGTAAACAAGTCGAAGTTAAAAGAGCTCTTCCTAAAGATGCTAACCCTGGAA
TAGCCAGTGGTGGTGGTCGTGGCAGTGGTGGAGCTGGAGGGTTTCCGGGCTATGGTGGTTCTGG
TGGAAGTGGCTATGAGGGTCGTGTGGATTCTAATAGATACATGCAGCCGCAAAACACTGGAAGT
GGTTATCCTCCTTATGGTGGTTCTGGGTATGGTACTGGTTATGGTTATGGAAGCAATGGTGTAG
GTTATGGGGGTTTTGGTGGGTATGGCAATCCAGCTGGTGCGCCTTATGGGAATCCTAGTGTCCC
TGGAGCTGGGTTTGGAAGTGGTCCAAGAAGTTCATGGGGCGCTCAAGCACCATCGGGTTATGGG
AATGTGGGATATGGAAATGCAGCTCCGTGGGGTGGTTCTGGTGGTCCTGGTTCAGCAGTAATGG
GTCAAGCTGGTGCATCTGCAGGTTATGGCAGTCAAGGTTATGGCTATGGTGGAAATGATTCCTC
TTACGGGACTCCATCTGCCTATGGTGCAGTAGGGGGGCGATCTGGGAATATGCCTAACAACCAT
GGTGGCGGTGGCTATGCGGATGCTTTAGATGGCTCTGGAGGCTATGGGAATCACCAAGGGAACA
ACGGGCAAGCTGGTTATGGTGGAGGTTATGGAAGTGGTAGGCAAGCTCAACAACAGTGAttgaa
gaagaaatactactagaatgtggttttatcgctgaccttgaaacctcctgctttccgccttaac
catgtcacgtctttggcggttagaccaggaggtggacctacgctggattatctcttttgttagt
ttctcaataagttgttttcaggcaattccggatactatttcctatcaagttgtagttttaagt
ttgcgtgcttatttatatttgtcgctttggaatggttttctttctctgttatcctctagtgttt
gtgtttaacgatacatcctccagattatcattattcatctccctttggttcattcatttttgt
tgaatattccattcacagattcttgcttttgcatctcctctgtttaggggaagatgatttgctc
agtgttcaatgtgatctaagaaaagtgtttggtagagcaagagctgcaataaatcactttgaga
ttgcgttgttacatgaaggtcgtgttggcggaaacttaacagtccca

seq id no 25: At2g33410 protein sequence

MESDQGKLFIGGISWDTDENLLREYFSNFGEVLQVTVMREKATGRPRGFGFVAFSDPAVIDRVL
QDKHHIDNRDVDVKRAMSREEQSPAGRSGTFNASRNFDSGANVRTKKIFVGGLPPALTSDEFRA
YFETYGPVSDAVIMIDQTTQRPRGFGFVSFDSEDSVDLVLHKTFHDLNGKQVEVKRALPKDANP
GIASGGGRGSGGAGGFPGYGGSGGSGYEGRVDSNRYMQPQNTGSGYPPYGGSGYGTGYGYGSNG
VGYGGFGGYGNPAGAPYGNPSVPGAGFGSGPRSSWGAQAPSGYGNVGYGNAAPWGGSGGPGSAV
MGQAGASAGYGSQGYGYGGNDSSYGTPSAYGAVGGRSGNMPNNHGGGGYADALDGSGGYGNHQG
NNGQAGYGGGYGSGRQAQQQ

FIGURE 6

seq id no 26: At5g47620 (contig mnj7 - coordinates: 62491-62504, 62602-63661, 64000-64221)

tgagcattgcttatttgcttccatccatttttgttccttttaattcgatttggattgcagaaaa
agaaaagaaaagaaaagactaaaaatttggacgataagcagaaaagagagaggagggcctctcg
ccctcttattaaaaccttgccttctccaaatctgaagatttctcaatcctaaaatctttttttt
ttcctctttctccgtttctttatttcggtattacacacatacatagattctctgtcttctggg
tttttcattccttccttcctccaagcttacacctttattgatcatttgtgtttttttttgtttc
tgcaggaatccaagatcgtgggtcgatcggttttacacaatccgatcacgacccatctgctct
ttttcatcctattttgcttcccttgaggtgtttctatcgattccattctccttctcacttagat
cgatatagaatctggaaccaaaaacaaacctttttttgtttgtttggcagaaATGGAAATGGAA
TCATGTAAGCTCTTCATCGGTGGTATATCTTGGGAAACCAGTGAAGATCGTCTTCGTGACTATT
TTCACAGTTTTGGTGAGGTTTTAGAGGCTGTTATTATGAAGGATCGTGCCACTGGCCGTGCTCG
TGGCTTTGGTTTCGTTGTCTTTGCTGATCCTAATGTTGCTGAAAGAGTCGTCTTGCTTAAACAT
ATCATTGATGGTAAAATTgtaagtttcctcctgctatataccaacatacattgcttccaatttc
aacaatcttcctgcttacttgcttcattttgaggttgctgcttctcaaagcaaagcaaagctac
tcacttttattccttcctgttttagttagtagactctattgtttacaatcagctttgccgctct
gataaatgcatatctttgtcagaagttgttcatttcacactcacaaataaaaatgtaaaacttg
gatcgtttcatatcctcatgtgaaagaaagtggttcacaatgaatgaaaaactgctttctttga
gttgtgtcgtgtgttgattttctccatgatatacagGTTGAGGCAAAGAAGGCTGTTCCAAGAG
ATGATCACGTAGTATTTAATAAAAGTAACAGCAGCCTTCAGGGATCACCTGGCCCATCAAACTC
CAAGAAGATCTTTGTGGGAGGTTTGGCATCATCCGTGACAGAGGCTGAGTTCAAAAAGTATTTT
GCTCAGTTTGGGATGATCACTGATGTTGTGGTGATGTATGACCACAGAACCCAGCGGCCTAGAG
GCTTTGGGTTCATTTCATATGACTCTGAGGAAGCTGTTGACAAAGTACTGCAGAAGACATTCCA
CGAACTCAATGGTAAGATGGTGGAGGTCAAACTGGCTGTTCCTAAGGATATGGCTCTCAACACA
ATGCGGAACCAAATGAATGTAAATAGCTTTGGCACTAGTAGAATCAGTTCATTACTGAATGAGT
ACACCCAGGGATTCAGCCCGAGTCCAATCTCTGGTTATGGAGTGAAACCTGAAGTTAGGTACAG
TCCAGCAGTAGGTAATAGGGGAGGATTCTCACCGTTTGGACATGGATACGGAATCGAGCTGAAT
TTTGAGCCAAACCAGACTCAGAACTACGGTTCTGGTTCCAGTGGAGGCTTTGGACGACCCTTTA
GCCCTGGATATGCTGCGAGTCTCGGCAGGTTCGGTAGCCAAATGGAGTCGGGAGGAGCTAGTGT
TGGGAACGGTTCTGTCCTAAATGCAGCACCAAAGAACCATTTATGGGGAAATGGTGGTCTAGGT
TACATGTCAAACTCTCCGATATCAAGAAGCAGCTTCAGTGGAAACTCTGGAATGTCTTCACTAG
GCAGCATTGGTGACAACTGGGGAACAGTTGCACGTGCACGCAGTAGCTACCACGGTGAGAGAGG
AGGTGTAGGATTAGAAGCAATGAGAGGAGTTCATGTTGGTGGTTACAGCAGCGGCTCAAGCATC
TTGGAGGCAGACTCTCTGTACAGCGACTCGATGTGGCTTTCGCTGCCTGCAAAGGCAGAGGAAG
GATTGGGAATGGGACCATTGGACTTCATGTCTAGAGGACCAGCTGGATACATCAACAGGCAACC
AAACGGAGgtatgaataatgaatgaatgaacgccttttttctatccgagaattcaagcatttgt
agaaaatctgatgatatcatatgaaaatggtgttgttgcagGAATTGCAGCTTAGagaagtgac
aaatctataccatggagatcagatgattgcagaagagagtttttagaagaggaaaaaagtttat
taaaaaaaaaaaaattattggtaccaaaaagcttaaagctttatttacttttactatttga
tttgttgttatagctttcttttcacccttttttctaatttggggttttgtttcttttgttttta
tcgttaaagaaaaaagatgtaaacttgagtgatataaaaagagacaaagaaacaatgaagtgta
ttttgttcttgtctttctctctcttttatcatctaaatccatatattgacaaattcaaacatga
aaacgaattaaaaaaagagcaatttgcctagaatgtaggcaacgtagtgtgaggacgacgtgtg
gcaaacatgtggatgatgataagccacaggacaaagaaagcaatccctcatccatcgcaataat
atccattaatgtgaagtggaccaaaagagagagaagcgagtgt

FIGURE 6

seq id no 27: At5g47620 protein sequence

MEMESCKLFIGGISWETSEDRLRDYFHSFGEVLEAVIMKDRATGRARGFGFVVFADPNVAERVV
LLKHIIDGKIVEAKKAVPRDDHVVFNKSNSSLQGSPGPSNSKKIFVGGLASSVTEAEFKKYFAQ
FGMITDVVVMYDHRTQRPRGFGFISYDSEEAVDKVLQKTFHELNGKMVEVKLAVPKDMALNTMR
NQMNVNSFGTSRISSLLNEYTQGFSPSPISGYGVKPEVRYSPAVGNRGGFSPFGHGYGIELNFE
PNQTQNYGSGSSGGFGRPFSPGYAASLGRFGSQMESGGASVGNGSVLNAAPKNHLWGNGGLGYM
SNSPISRSSFSGNSGMSSLGSIGDNWGTVARARSSYHGERGGVGLEAMRGVHVGGYSSGSSILE
ADSLYSDSMWLSLPAKAEEGLGMGPLDFMSRGPAGYINRQPNGGIAA

seq id no 28: NM_196957 DNA sequence (*Oryza sativa*)

atggagtcggatcaggggaagctgttcatcggcggcatctcgtgggagaccaccgaggagaagc
tccgcgaccacttcgccgcctacggcgacgtctcccaggccgccgtcatgcgcgacaagctcac
cggccgccccgcggcttcggcttcgtcgtcttctccgacccttcctccgtcgacgccgccctc
gtcgacccccacaccctcgacggccgcacggttgatgtgaagcgggcgctctcgcgggaggagc
agcaggccgcgaaggcggcgaaccctagcgcgggggggaggcacgcctccggtgggggcggtgg
tgggggaggcgccggtggtggtggtggtggcggcggtggtgacgccggcggtgcgcggacgaag
aagatcttcgtcggcgggctgccctccaacctgacggaggacgagttccggcagtacttccaga
cctacggggtcgtcaccgacgtcgtcgtcatgtacgaccagaacacgcagcggccgaggggttt
cgggttcatcaccttcgacgcggaggacgccgttgaccgcgtgctgcacaagaccttccatgac
ctgagcgggaagatggtggaggtgaagcgcgccctgcccagggaggccaaccctggctccggca
gtggtggccgttccatgggaggcggcggtgggggttaccagagtaacaatgggccgaactccaa
ttctgggggctatgatagcagaggtgacgctagcaggtatggtcaggcgcagcagggtagtggt
ggttatcccggttatggtgctggaggatatggtgctggtacggttggttatggatatgggcatg
ctaaccctggaactgcgtatgggaattatggggctggaggatttggaggtgttcctgctgggta
tggtgggcattatggcaatccaaatgcgcctggttcaggttaccagggtggtcctccaggagca
aacagaggaccatggggtggtcaagctccgtctggttatggcactgggagttatggtggcaatg
caggctatgctgcttggaacaactcttctgctggaggtaatgcacccactagtcaggccgctgg
tgcaggcacaggctatgggagccagggctatggatatggtggatatggaggagatgcatcgtat
ggtaatcatggtggatatgggggttatggaggaaggggagatggtgctggcaatccagctgctg
gcggtggatctgggtatggtgctggctatggaagcgggaatggcggttctggttatccaaatgc
ttgggctgatccttcacaaggtggagggtttggggcttcagtcaatggagtgtctgaaggccaa
tcaaattatggcagtggttatggtggtgtgcaacctagggttgctcagtaa

seq id no 29: NP_921939 protein sequence (*Oryza sativa*)

MESDQGKLFIGGISWETTEEKLRDHFAAYGDVSQAAVMRDKLTGRPRGFGFVVFSDPSSVDAAL
VDPHTLDGRTVDVKRALSREEQQAAKAANPSAGGRHASGGGGGGGGAGGGGGGGGDAGGARTK
KIFVGGLPSNLTEDEFRQYFQTYGVVTDVVVMYDQNTQRPRGFGFITFDAEDAVDRVLHKTFHD
LSGKMVEVKRALPREANPGSGSGGRSMGGGGGGYQSNNGPNSNSGGYDSRGDASRYGQAQQGSG
GYPGYGAGGYGAGTVGYGYGHANPGTAYGNYGAGGFGGVPAGYGGHYGNPNAPGSGYQGGPPGA
NRGPWGGQAPSGYGTGSYGGNAGYAAWNNSSAGGNAPTSQAAGAGTGYGSQGYGYGGYGGDASY
GNHGGYGGYGGRGDGAGNPAAGGGSGYGAGYGSGNGGSGYPNAWADPSQGGGFGASVNGVSEGQ
SNYGSGYGGVQPRVAQ

FIGURE 6

seq id no 30: AK067725 DNA sequence (*Oryza sativa*)

ggtccattatttataccatttccgcgtccccccaccctcctcccccgctttcccaatcgaggcg
agcaccgcaattgcagggttccggaggccgaataaaaaagtttggcctctccccgcaaaaaagt
aaaaaacccaaaacaaccatccaccagcgcatcgcggcaccgcgagcgagcgagcggagggagg
gaggtggagagcaaaagttcgataaaaggagaggaggagacgaagcgtcgaagcccaagtaaca
tccccccaacctccgcctcctcctcctcccccctcctcccatgcccgcatcgagatcttagccgc
gccggagatcgagagggaggagcggcgacgcgggcgcccccgatccctcctcctcgccgccgcc
gccgccggcggcgccggagcagcagcagccgacgacgacgacgaccgccgcagcagccgatcgg
gggaggaggggaggggaggacgcgatggaggcggactccgggaagctcttcgtcggcggcatct
cgtgggagacggacgaggaccgcctccgcgagtacttcagccggttcggggaggtcaccgaggc
cgtcatcatgcgggaccgcaacaccggccgcgcccgtgggttcggcttcgtggtcttcaccgac
gcaggcgtcgccgagcgggtcaccatggataagcacatgatcgacgggcgcatggtggaagcga
agaaagctgttcccagggacgaccagagcatcaccagcaagaacaatggcagcagcatagggtc
acctggaccaggccgtactagaaagatctttgttggaggcttggcctctaatgttactgaggtt
gaatttagaaggtattttgagcaatttggtgtgattacggatgtggttgtcatgtacgaccaca
acacgcagaggcctaggggctttggattcatcacctatgactcagaagatgcggtggacaaggc
actgcacaagaacttccatgagctgaatggtaagatggttgaggtcaagagagctgttccaaag
gagcaatcacctggacctgctgcacgttcacctgcgggagggcagaactatgctatgagcaggg
tccatagcttcttgaatggtttcaaccagggttataacccaaaccctattggaggttatggcat
gagggttgatggaaggtatggtctgcttacaggcgcacggaatggattctcttcatttggccct
ggttatggaatgggcatgaattctgaatctgggatgaatgcgaattttggcgccaattctagtt
ttgtcaataactccaatgggcggcagataggttcattctacaatggtagttcaaacagattagg
tagtcctattggttatgttggtcttaatgatgattcaggatcactattgagttcaatgtcaagg
aatgtttggggtaatgaaaatctgaactacccaaacaaccccacaaacatgagttcttttgcac
catctggaactggaggtcaaatgggtattaccagtgacggtattaattggggagggcctactcc
tggccatggaatgggcaacatttcaagccttgggctggctaaccttggccgtggagctggagac
agttttggcttgccttctggcagctatggaaggagcaatgcaactggtaccattggtgaaccct
tctctgcaccacccaatgcatatgaagtgaacaatgcagatacatatggcagcagctccattta
tggagactcaacttggaggttcacgtcatctgagattgatatgcctccttttggtaatgacctt
ggaaatgttgatccagatatcaaatcaaacataccagcaagttacatgggcaactatactgtta
ataataatcagacaagcagaggtatcacttcctagcgagagtactattatattcatatatgact
tgggatagatgaaagaagcattatatcaggtattcaggtgcatgactatgaattggtgatatca
ggttaatatacgggttagttaattgtttctagctaaccagaggtgtggtttatggacaccacca
tgctagaggagcgaatacaaacgttttgtgaaggtttcagattttagtttaattcctacatgta
ttaggtcttggtttttgaatgagatgtgcagtggtgattgcggcacatacttagagtgttccaa
cataagctggaatcctgtcatatggacaaacttgtataccaaaggaatgctttattatcttgcc
catttatggctacattagctcgcttgttttcattccctttttaaccaattccatttgtatacta
gagatctgcttgacttactagtgaaactattcggggacgccgatcctatctttgcagttggctc
ccagaaataaagccaccaaaagtgcatacttatttgttctaccttgatttgccatatgtatatg
cttctgttcgttttaaaatagaactttgggtttgatt

seq id no 31: AK067725 protein sequence (*Oryza sativa*)

MEADSGKLFVGGISWETDEDRLREYFSRFGEVTEAVIMRDRNTGRARGFGFVVFTDAGVAERVT
MDKHMIDGRMVEAKKAVPRDDQSITSKNNGSSIGSPGPGRTRKIFVGGLASNVTEVEFRRYFEQ
FGVITDVVVMYDHNTQRPRGFGFITYDSEDAVDKALHKNFHELNGKMVEVKRAVPKEQSPGPAA

FIGURE 6

rspaggqnyamsrvhsflngfnqgynpnpiggygmrvdgryglltgarngfssfgpgygmgmns
esgmnanfganssfvnnsngrqigsfyngssnrlgspigyvglnddsgsllssmsrnvwgnenl
nypnnptnmssfapsgtgggqmgitsdginwggptpghgmgnisslglanlgrgagdsfglpsgs
ygrsnatgtigepfsappnayevnnadtygsssiygdstwrftsseidmppfgndlgnvdpdik
snipasymgnytvnnnqtsrgits

seq id no 32: AK070544 DNA sequence (*Oryza sativa*)

ttggagatagaatagagagagacacacaaacacctacaacaccaacaacaacaagagaaagaga
gaaagaagagaaggaaaggagaggaagaagaggtggtggtggtggtggtggtgtgtggcct
ccttcccctccctcctctcgcgaggttgccatgcctcccccaagatcgatccaacccgatcatc
aatcggggcggggaaggaggaggaggggatggaggcggacgccgggaagctgttcatcggcggc
atctcgtgggacaccaacgaggaccgcctccgcgagtacttcgacaagtacggcgaggtggtgg
aggccgtcatcatgcgcgaccgcgccaccggccgcgcccggggattcggcttcatcgtcttcgc
tgaccctgccgtcgccgagcgggtcattatggagaagcacatgatcgatggccgcatggtggag
gcgaagaaagctgttcccagggacgatcagcacgctcttagcaagagcggcgggagcgctcatg
gatcgccggggcccagccgcaccaagaagatattcgttgggggctagcgtccaccgtgacgga
ggcggacttcaggaagtactttgagcagttcgggacgatcaccgatgtcgtggtgatgtatgat
cacaacacgcagcgtcccagaggttttgggttcattacgtacgattcggaggatgctgtggaca
aggcattgttcaagaccttccatgaactgaacggtaagatggttgaggtcaagcgcgcggttcc
taaggaactatcacctgggcctagcatgcgttctcctgtcggtggattcaactatgccgtgaac
agagccaataacttcctcaatggatacacccagggttataatccgagcccagtcggtggctatg
gaatgaggatggatgcaaggtttgggcttctatcgggtggccgtagtagttatccttcttttgg
tggtggttatggagtcggtatgaattttgatccagggatgaaccctgctattgggggaagctca
agcttcaacaacagtctccagtatggaaggcagcttaatccatactacagtggaaattctggta
gatacaatagcaatgttagctatggtggagtcaatgacagtactgggtcagtgttcaactcgct
ggctcgtaatttatggggtaattcaggtcttagttactcttccaactctgcaagctctaattcc
ttcatgtcatctgccaatgggggccttggtggaattgggaacaacaatgtgaattggggaaacc
ctcctgtgcctgcacaaggtgctaatgctggcccaggctatggcagtgggaacttcggttatgg
atccagtgaaaccaactttggtctcggtaccaatgcttatggaaggaatgctggatctggtgtt
gttaatacattcaatcaatcaaccaatgggtatggaaggaactttggagattcatcaggaggag
gtggcggtggtggcggtggctccatctatggagacacaacttggagatccggatcttctgagct
tgatggaaccagcccatttggctatgggcttgggaatgcagcttcagatgttacagcaaagaac
tcagcaggttacatggggcattaacaaatagagcaatgtcgccgcctaggaatctttttcacat
acaacatttgtcaaaataggttgaggagagaaccacaggtgcatcaggtgcaaattttgaacct
cacatgatttacagaaatgggttagttaatagagctaaccaccagggatttggtcaatgagatc
agatatatatcctcagagaaccatttaaacgtatttccattttatgtaaggtttgagattgtgg
tttcggatttctacagcgagtttaggttttggcaaccttgtgtttttttcttggttgagatgtga
agtaagattgcgggatatatatatctgaagagtgttcagttgtacggcggcgctgcccccatat
aggcccccctttttgggtttttgttcttatagtagaaactgctctagcgttttgcaaattgtgt
gctagctgttgttatcaggatgataatttttttcccctttcttggtttttatcttactgaagtgt
atgtaccagagatcttgctggtctgtgtttttcctagtggaacttttgagggatgccccttctg
ggtctcaaagaataataatgctacattatattctaattcattttgaggctttctaaggctatat
attatttgtatgtaccctgctggaacatctgtacattctgatgctctttgcaatttgcctttgt
gctgcttttgc

FIGURE 6

seq id no 33: AK070544 protein sequence (*Oryza sativa*)

MEADAGKLFIGGISWDTNEDRLREYFDKYGEVVEAVIMRDRATGRARGFGFIVFADPAVAERVI
MEKHMIDGRMVEAKKAVPRDDQHALSKSGGSAHGSPGPSRTKKIFVGGLASTVTEADFRKYFEQ
FGTITDVVVMYDHNTQRPRGFGFITYDSEDAVDKALFKTFHELNGKMVEVKRAVPKELSPGPSM
RSPVGGFNYAVNRANNFLNGYTQGYNPSPVGGYGMRMDARFGLLSGGRSSYPSFGGGYGVGMNF
DPGMNPAIGGSSSFNNSLQYGRQLNPYYSGNSGRYNSNVSYGGVNDSTGSVFNSLARNLWGNSG
LSYSSNSASSNSFMSSANGGLGGIGNNNVNWGNPPVPAQGANAGPGYGSGNFGYGSSETNFGLG
TNAYGRNAGSGVVNTFNQSTNGYGRNFGDSSGGGGGGGGSIYGDTTWRSGSSELDGTSPFGYG
LGNAASDVTAKNSAGYMGH

seq id no 34: CK210974 DNA sequence (*Triticum aestivum*)

AAAAAGCAGGTGGGACCGGCCCGGAATTCTCGGGATATCGTCGACCCACGCGTCCGCGCACCCG
AGCGCGAGAGAATCCGAGGAGAGGAGCGGCGCAAGGAGGCGGTGATGGAGTCGGATCAGGGCAA
GCTCTTCATCGGCGGCATCTCCTGGGAGACGACGGAGGAGAAGCTGCAGGAGCACTTCTCCAAC
TTCGGCGAGGTCTCCCAGGCCGCCGTCATGCGCGACAAGCTCACTGGCCGCCCGCGGGGCTTCG
GCTTCGTAGTCTACGCCGACCCCGCCGCCGTCGACGCCGCCCTCCAGGAGCCCCACACCCTCGA
CGGCCGCACGGTCGATGTGAAGCGGGCGCTCTCGCGGGAGGAGCAGCAGGCTACCAAGGCGGTG
AACCCTAGCGCAGGAAGGAACGCTGGAGGTGGTGGCGGCGGCGGCGGCGGCGGCGGCGATGCCG
GTGGTGCTAGGACAAAGAAGATTTTTGTGGGCGGACTGCCCTCCAGTCTGACAGATGAGGAGTT
CCGGCAGTACTTCCAGACCTTCGGGGCTGTCACCGATGTTGTGGTGATGTATGACCAGACAACA
CAGCGTCCCCGGGGCTTCGGCTTCATTACCTTTGACTCGGAGGATGCGGTTGACCGTGTGCTGC
ACAAAACCTTCCACGATCTTGGAGGGAAGATGGTAGAGGTGAAGCGTGCTCTGCCCCGAGAGGC
GAATCCTGGCTCTGGCGGCGGCGGCCGTTCCATGGGAGGTGGGGGGTTTCATAGTAACAATGGA
CCCCACTCCAATGCTAGCAGCTATGATGGCAGAGGCGATGCTAGCAGATATGGGCAGGCGCAGC
AAGGCATGGGTGGCTACCCAGGTTATGGTGCTGGAGCTTATGGCAGTGCTCCAACTGGGTTTGG
ATATGGGCCACCCAATCCGGGAACTACTTATGGAAATATTGGGTCTGCAGGGTTAGGAGCTTTT
CCTTGGTGCGTATGCGGGGGGCTTATGGGCAACCCAGGTGGCTGCGGGTTTCGGGTTACCCGGG
GGGGCCCCTCCGGGGCCCTAAATAAGGGACCCTGGGGGCAGCCAAACCTCCGCCCTGGTTTATG
GCACCTGGGGGCTTTATCCTGGGCACGTGCGGGGCTATTGGGTGCGTGGAAATAACCC

seq id no 35: CK210974 protein sequence (*Triticum aestivum*)

MESDQGKLFIGGISWETTEEKLQEHFSNFGEVSQAAVMRDKLTGRPRGFGFVVYADPAAVDAAL
QEPHTLDGRTVDVKRALSREEQQATKAVNPSAGRNAGGGGGGGGGGDAGGARTKKIFVGGLPS
SLTDEEFRQYFQTFGAVTDVVVMYDQTTQRPRGFGFITFDSEDAVDRVLHKTFHDLGGKMVEVK
RALPREANPGSGGGGRSMGGGGFHSNNGPHSNASSYDGRGDASRYGQAQQGMGGYPGYGAGAYG
SAPTGFGYGPPNPGTTYGNIGSAGLGAFPWCVCGGLMGNPGGCGFRVTRGGPSGALNKGPWGQP
NLRPGLWHLGALSWARAGLLGAWK

seq id no 36: CA124210 DNA sequence (*Saccharum officinarum*)

AGAATTCNCGGTTCGACCTACGCGTCCGCCCGGAATCCCCAATTCCGCTCTCTTCCTCTCTCCC
TCTCTCCCCCACCGCAGCATCAGGCGAGCGCGAGGCGGAGGTGGAGGAGAGATGGAGTTGGACC
AGGGCAAGCTCTTCATCGGCGGCATCTCCTGGGAGACGACGGAGGAGAAGCTGAGCGAGCACTT
CTCCGCCTACGGCGAGGTTACGCAGGCCGCCGTCATGCGGGACAAGATCACCGGCCGCCCCCGT
GGCTTCGGGTTCGTCGTCTTCGCCGACCCCGCCGTCGTCGACCGAGCGCTGCAGGACCCCCACA
CCCTCGACGGCCGCACGGTCGATGTGAAGCGGGCACTCTCGCGGGAGGAGCAGCAGGCCTNCAA

FIGURE 6

GGCCGCGAACCCTAGCGGTGGGAGGAACACTGGCGGTGGANGANGCGGCGGGTGGCGGGGCGGC
GATGCAAGTGGTGCTCGGACCCAGGAAGATCTNTGGGGGGCCGGCTTGCCTTCTACTCTGACTG
ANGGATGGGTTTCGGCAGTACTTTCCGGACCTTCGGAGGGGTCACTGATGGTTGGTGGCCATGG
TTGAACCGGAACAAGCAATTGCCCGCGTTGGTTTTGGAATCAATACTTTTGAACTTTAAGATTC
CGGTGAACCGCTGCTGGCCAAGAACTTTCATGACCTGGTGGGAAGATGGTTTAAGGTGAACCAG
CATTGCGCCCTTGAGGCGAACCCTGGGGGTTCTGGAACGGGCCGTTCTGGGGGAAATGGGGGCT
TTCTAGCAACCATGGCCTTACCCCCGTTTTGG

seq id no 37: CA124210 protein sequence (*Saccharum officinarum*)
MELDQGKLFIGGISWETTEEKLSEHFSAYGEVTQAAVMRDKITGRPRGFGFVVFADPAVVDRAL
QDPHTLDGRTVDVKRALSREEQQAXKAANPSGGRNTGGGXXGGWRGGDASGARTQEDLWGAGLP
STLTXGWVSAVLSGPSEGSLMVGGHG

seq id no 38: beta-expansin promoter (*Oryza sativa*)
aaaaccaccgagggacctgatctgcaccggttttgatagttgagggacccgttgtgtctggttt
tccgatcgagggacgaaaatcggattcggtgtaaagttaagggacctcagatgaacttattccg
gagcatgattgggaagggaggacataaggcccatgtcgcatgtgtttggacggtccagatctcc
agatcactcagcaggatcggccgcgttcgcgtagcacccgcggtttgattcggcttcccgcaag
gcggcggccggtggccgtgccgccgtagcttccgccggaagcgagcacgccgccgccgccgacc
cggctctgcgtttgcaccgccttgcacgcgatacatcgggatagatagctactactctctccgt
ttcacaatgtaaatcattctactattttccacattcatattgatgttaatgaatatagacatat
atatctatttagattcattaacatcaatatgaatgtaggaaatgctagaatgacttacattgtg
aattgtgaaatggacgaagtacctacgatggatggatgcaggatcatgaaagaattaatgcaag
atcgtatctgccgcatgcaaaatcttactaattgcgctgcatatatgcatgacagcctgcatgc
gggcgtgtaagcgtgttcatccattaggaagtaaccttgtcattacttataccagtactacata
ctatatagtattgatttcatgagcaaatctacaaaactggaaagcaataagaaatacgggactg
gaaaagactcaacattaatcaccaaatatttcgccttctccagcagaatatatatctctccatc
ttgatcactgtacacactgacagtgtacgcataaacgcagcagccagcttaactgtcgtctcac
cgtcgcacactggccttccatctcaggctagctttctcagccacccatcgtacatgtcaactcg
gcgcgcgcacaggcacaaattacgtacaaaacgcatgaccaaatcaaaaccaccggagaagaat
cgctcccgcgcgcggcggcgacgcgcacgtacgaacgcacgcacgcacgcccaaccccacgaca
cgatcgcgcgcgacgccggcgacaccggccgtccacccgcgccctcacctcgccgactataaat
acgtaggcatctgcttgatcttgtcatccatctcaccaccaaaaaaaaaaggaaaaaaaaacaa
aacacaccaagccaaataaaagcgaca

seq id no 39: Artificial sequence (primer prm00405)
ggggacaagtttgtacaaaaaagcaggcttcacaatggattatgatcggtacaagttat

seq id no 40: Artificial sequence (primer prm00406)
ggggaccactttgtacaagaaagctgggtttaaaagagtccaaagaatttcact

FIGURE 6